THÈSE

DE DOCTORAT D'ÉTAT ÈS SCIENCES MATHÉMATIQUES

présentée

A L'UNIVERSITÉ PARIS VI

pour l'obtention

DU GRADE DE DOCTEUR ÈS SCIENCES

par

Kada ALLAB

Sujet de la Thèse :

APPROXIMATION D'OPÉRATEURS NON LINÉAIRES ET IDENTIFICATION DES SYSTÈMES DÉTERMINISTES OU ALÉATOIRES.

Soutenue le 28 Mai 1973, devant la Commission d'Examen

MM. J. VILLE — *Président*
R. FORTET
J L. LIONS
Y. CHERRUAULT — *Examinateurs*

THÈSE

DE DOCTORAT D'ÉTAT ÈS SCIENCES MATHÉMATIQUES

présentée

A L'UNIVERSITÉ PARIS VI

pour l'obtention

DU GRADE DE DOCTEUR ÈS SCIENCES

par

Kada ALLAB

Sujet de la Thèse :

APPROXIMATION D'OPÉRATEURS NON LINÉAIRES ET IDENTIFICATION DES SYSTÈMES DÉTERMINISTES OU ALÉATOIRES.

Soutenue le 28 Mai 1973, devant la Commission d'Examen

MM. J. VILLE — *Président*
R. FORTET
J L. LIONS
Y. CHERRUAULT — *Examinateurs*

UNIV.
LEIDEN
BIBL.

AVANT-PROPOS

Je remercie vivement M. le Professeur R. FORTET dont l'enseignement et les nombreux conseils m'ont beaucoup aidé à mener à bien ce travail.

J'adresse également mes vifs remerciements à M. le Professeur J. VILLE pour m'avoir fait l'honneur de présider le jury de cette thèse et MM. les Professeurs J.L. LIONS et Y. CHERRUAULT d'avoir bien voulu participer à ce jury.

I N T R O D U C T I O N

D'un point de vue physique un système peut grossièrement être considéré comme un mécanisme faisant correspondre à une action (on dira une entrée) une réaction (une sortie).

La connaissance d'un système revient à celle des lois qui régissent son comportement.

L'étude du comportement à partir des lois élémentaires est théoriquement possible, mais en pratique elle est souvent impossible si le système est complexe, si les phénomènes présents ne sont pas, ou sont mal connus etc...

Ceci nous amène à rechercher une approximation du comportement du système dans l'ignorance totale des lois qui le régissent et des phénomènes présents. Cette approximation ou identification du système doit se faire seulement à partir de certaines entrées et des sorties correspondantes.

D'un point de vue mathématique un système peut être représenté par un opérateur, celui-ci faisant correspondre à une fonction (la fonction entrée), une autre fonction (la fonction sortie). L'identification du système revient à la détermination d'une approximation de l'opérateur.

La façon d'aborder le problème d'identification dépend étroitement des connaissances " à priori" qu'on possède sur celui-ci.

On a d'abord cherché à représenter des systèmes physiques linéaires et de forme bien connue; cela a fourni les méthodes basées sur l'analyse indicielle et harmonique et sur la transformation de Laplace. Dans la période actuelle où l'identification cherche à représenter des systèmes plus compliqués, de forme moins bien connue mais toujours linéaires, les méthodes les plus étudiées sont celles reposant sur le modèle d'état :

$$\frac{dx}{dt} = Ax + Bu$$

$$y = Cx + Du$$

Cette approche est maintenant assez bien connue tant du point de vue pratique que du point de vue théorique. Elle présente en outre l'avantage de permettre la commande du système, la méthode de commande optimale correspondant à ce modèle d'état étant bien définie.

Dans le cas non linéaire néanmoins, aucune voie générale ne semble s'imposer actuellement. Les méthodes publiées sont plutôt spécifiques des problèmes particuliers étudiés, soit qu'elles utilisent des connaissances "à priori", soit qu'elles introduisent des non linéarités de forme particulière dans un modèle d'état etc...

Le travail présenté ici fournit une approche générale et systématique au problème de l'identification non linéaire. Le modèle obtenu permet de représenter les systèmes non linéaires déterministes ou stochastiques.

On a souvent cherché à représenter des opérateurs par des intégrales à noyau (théorème de Riesz pour les fonctionnelles linéaires continues par exemple, ou le théorème des noyaux pour les fonctionnelles bilinéaires).

On va montrer dans cette étude qu'un opérateur non linéaire peut être représenté par une combinaison linéaire d'intégrales à noyaux.

La première partie, relative à la recherche du modèle, étudie les propriétés des opérateurs polynomes et spéciallement ceux du type Hilbert Schmidt (on notera H-S; le chapitre 0 rappelle les principales définitions). Elle montre leur utilisation pour approximer des opérateurs non linéaires.

On montrera d'abord que tout opérateur non linéaire défini et continu sur un compact X de $L^2(T)$ (T intervalle réèl) peut être représenté par un polynome H-S (donc par des intégrales à noyaux); autrement dit l'ensemble des polynomes H-S est dense dans $C[X, L^2(T)]$.
Si au lieu de $L^2(T)$, l'opérateur est défini sur un Hilbert séparable, on verra qu'il peut être représenté par un opérateur polynome.

On étudiera ensuite la classe des opérateurs non linéaires analytiques dans $L^2(T)$; on montrera qu'ils admettent une représentation par une série d'opérateurs H-S. Le résultat sera d'abord établi pour les fonctionnelles, et moyennant conditions, sera étendu aux opérateurs dans $L^2(T)$.

En vue de son utilisation pour la représentation des transformations non linéaires d'un processus stochastique, on montrera que l'ensemble des polynômes H-S est dense dans $L^p[X,\mu]$ où X est un compact de $L^2(T)$ et μ une mesure sur X (définie par un processus par exemple; une discussion sera faite sur les transformations d'un processus aléatoire lorsque celui-ci est défini par un système projectif de mesures).

On discutera de l'utilisation de polynomes H-S orthogonaux pour représenter un système, de la représentation de systèmes à plusieurs entrées et celle des systèmes aléatoires.

Aprés le problème de la représentation, on passera à celui de l'entrée déterminante qui fixera les paramètres du modèle.

Bien que la solution pratique consiste à utiliser un bruit blanc pour identifier le système, on étudiera le problème d'identification par entrées déterministes. Ce problème sera approché par la notion de classes déterminables : étant donnée une fonction déterministe entrée, quels sont les systèmes identifiables à partir de cette fonction entrée seulement? En utilisant des méthodes d'analyse spectrale, on explicitera la classe des opérateurs qu'elle détermine.

La phase finale est celle de l'identification numérique. Le modèle H-S est utilisé pour représenter des systèmes à entrées déterministes ou aléatoires; l'identification se ramène à la détermination des noyaux du modèle. En utilisant le critère des moindres carrés, on montre que la résolution directe du problème conduit à un système d'équations intégrales.

Une méthode de résolution approchée est décrite; elle se ramène à la détermination de coéfficients (composantes des noyaux selon une base fixée) qui sont les solutions d'un système d'équations algébriques linéaires.

Ce système peut être résolu par les méthodes usuelles si les caractéristiques statistiques des entrées-sorties sont connues; si elles sont inconnues, on montrera comment l'algorithme de l'approximation stochastique permet d'obtenir la solution.

Le principe de l'approximation stochastique est décrit à la fin ainsi qu'une méthode d'accélération de la convergence de son algorithme.

La dernière partie est consacrée à la simulation. Un H-S d'ordre 2 est utilisé pour représenter un système; des algorithmes pratiques sont détaillés et appliqués à l'identification d'un opérateur non linéaire. Les résultats numériques de l'exemple sont présentés à la fin.

CHAPITRE 0

OPERATEURS DE HILBERT-SCHMIDT

ET OPERATEURS POLYNOMES

Ce chapitre se compose de rappels

1. Opérateurs de Hilbert-Schmidt :

Opérateurs compacts :

Soient H_1 et H_2 deux espaces de Hilbert; un opérateur $A \in \mathcal{L}(H_1, H_2)$ est compact s'il transforme toute partie bornée de H_1 en partie relativement compact de H_2 (partie dont l'adhérence est compacte).

Rappelons que si A est compact, il diffère d'un opérateur compact défini positif ($\langle Af, f\rangle > 0 \ \forall f \in H_1$) par une isométrie U ($\|Uf\| = \|f\|$). autrement dit, A=UT où T est un opérateur compact défini positif dans H_1, et U une isométrie appliquant l'image de T dans H_2.

Si H_1 et H_2 sont séparables, et si A est compact, il admet la représentation :

$$Af = \sum_{n=1}^{\infty} \lambda_n \langle f, e_n \rangle h_n \qquad (1)$$

où (e_n) et (h_n) sont des ensembles orthonormés dans H_1 et H_2, $\lambda_1, \lambda_2, \ldots$ étant des nombres positifs tendant vers 0 lorsque $n \to \infty$.

Inversement, tout opérateur ayant la forme (1), avec les mêmes conditions pour λ_n, e_n, h_n, est compact (les λ_n sont les valeurs propres de l'opérateur U de la décomposition A=UT, les e_n les vecteurs propres de T, et les h_n de la forme $h_n = Ue_n$). [1]

Opérateurs de Hilbert-Schmidt :

Un opérateur compact A=UT est dit opérateur Hilbert-Schmidt (on notera H-S) si :

$$\sum_{n=1}^{\infty} \lambda_n^2 < +\infty$$

λ_n étant les valeurs propres de T.

Soit A un opérateur H-S de H_1 dans H_2, tel que la série $\sum_1^\infty \|Af_n\|^2$ converge pour une base orthonormée $f_1, f_2, ..$ dans H_1. Alors $\sum_1^\infty \|Ag_n\|^2$ converge pour toute base orthonormée (g_n) dans H_1 et on a :

$$\sum_{n=1}^{\infty} \|Af_n\|^2 = \sum_{n=1}^{\infty} \|Ag_n\|^2$$

Cette condition nécessaire est aussi suffisante pour que A soit H-S; autrement dit, un opérateur A est H-S si et seulement si la série $\sum_1^\infty \|Af_n\|^2$ converge pour au moins une base orthonormée (f_n) de H_1.

Cette dernière proposition étant plus commode à vérifier , est souvent prise comme définition d'un opérateur H-S.

Au cours de l'étude, on utilisera parfois la norme H-S :

Posons $\|A\|_2^2 = \sum_1^\infty \|Af_n\|^2$; cette expression satisfait tous les axiomes d'une norme; on l'appelle norme H-S de A et on note $|||A|||$; elle est finie pour et seulement pour les opérateurs H-S; la comparaison avec la norme du sup donne :

$$\|A\| = \sup_{\|f\| \leq 1} \|Af\| \leq |||A|||$$

L'ensemble $\mathcal{H}$ des opérateurs H-S est un espace vectoriel normé pour la norme HS

Ainsi, un opérateur A de H_1 dans H_2 est H-S si et seulement si il admet la représentation :

$$Af = \sum_{n=1}^{\infty} \lambda_n \langle f, e_n \rangle h_n$$

où (e_n) et (h_n) sont des ensembles orthonormés dans H_1 et H_2 respectivement, $\lambda_n > 0$ et tel que la série $\sum_1^\infty \lambda_n^2$ converge.

Représentation intégrale d'un opérateur H-S défini dans un L^2 :

- Supposons $H_1=H_2=L^2(a,b)$; soient (ϕ_n) un système orthonormé dans $L^2(a,b)$ et A un opérateur H-S défini dans cet espace.
A s'écrit :

$$Af(x)=\sum_1^\infty \lambda_n \langle f,\phi_n\rangle \phi_n(x) = \sum_1^\infty \lambda_n \left[\int_a^b f(y)\overline{\phi_n(y)}dy\right]\phi_n(y)$$

avec $\lambda_n \to 0$ et $\sum_n \lambda_n^2 < \infty$, conditions qui font que $\sum_1^\infty \lambda_n \phi_n(x)\overline{\phi_n(y)} = K(x,y)$ converge en norme, d'où en permutant Σ et $\int$:

$$Af(x) = \int_a^b K(x,y)f(y)dy$$

- D'une manière générale, si A est un opérateur H-S :

$$A : L^2(\Omega,\alpha,m) = L^2(\Omega) \to L^2(\Omega',\alpha',m') = L^2(\Omega')$$

une fonction $K \in L^2(\Omega\times\Omega')$ telle que :

$$(Af)(x) = \int_\Omega K(x,y)f(y)m(dy)$$

L'étude de certaines propriétés des opérateurs H-S dans un Hibert séparable se ramène à celle d'opérateurs H-S dans un L^2 :

Soit Ω un sous ensemble compact de R^n. Tout opérateur H-S A dans un Hilbert séparable H est unitairement équivalent à un opérateur H-S K dans l'espace $L^2(\Omega)$; en particulier il est équivalent à un opérateur H-S intégral dans $L^2[0,1]$. [2]

($\exists$ une application unitaire U: $H \to L^2(\Omega)$ telle que $K=UAU^{-1}$)

2. Opérateurs polynomes

définition :

Soient X et Y deux espaces vectoriels complexes; une application $y = P(x)$ de X dans Y définie pour tous les x est un opérateur polynome de degré m si :

$$P(x_1 + \alpha x_2) = \sum_{n=0}^{m} P_n(x_1, x_2)\alpha^n \qquad \forall x_1, x_2 \in X, \quad \alpha \text{ complexe}$$

$P_n(x_1,x_2)$ étant indépendants de α.

Le degré est exactement m si $P_m(x_1,x_2) \not\equiv \theta$, θ étant l'élément nul de Y.

Un polynome $P(x)$ est dit homogène de degré n si $P(\alpha x)=\alpha^n P(x)$.

Un polynome de degré m est la somme de polynomes homogènes $P_n(x)$ de degré n :

$$P(x) = \sum_{n=0}^{m} P_n(x)$$

On retrouve les propriétés algébriques des polynomes usuels : la somme de plusieurs polynomes est un polynome dont le degré n'excède pas le plus grand degré présent dans la somme; le produit de deux polynomes de degrés n et m en x est un polynome de degré n+m en x. Cependant un polynome n'est pas toujours continu (X et Y topologiques).

forme polaire d'un polynome :

La forme polaire d'un polynome homogène $P_m(x)$ de degré m est l'application P de X^m dans Y définie par : [3]

$$P(x_1,..,x_m) = \frac{1}{m!}\, \Delta^m_{x_1,..,x_m} P_m(a) \qquad \text{avec } \Delta^m_{h_1,..,h_m} = \Delta^{\nu_1}_{h_1} \ldots \Delta^{\nu}_{h_m}$$

$$\text{et } \Delta^{\nu}_{h} P_m(a) = \Sigma(-1)^{\nu-\mu} \binom{\nu}{\mu} P_m(a+\mu h)$$

($\nu^{\text{ème}}$ différence de P)

Proposition :

Si $P_n(x)$ est un polynome homogène de degré n, sa forme polaire est une application de X^n dans Y, n-linéaire et symétrique; elle est unique

et à l'aide de sa forme polaire P le polynome s'écrit :

$$P_n(x) = P(x,\ldots,x)$$

Inversement, si $F(x_1,..,x_n)$ est une application n-linéaire symétrique de X^n dans Y, alors $P_n(x)=F(x,..,x)$ est un polynome homogène de degré n et $F(x_1,..,x_n)$ est la forme polaire de $P_n(x)$.

Cette proposition permet, et c'est souvent plus commode, de considérer un polynome homogène comme la forme $F(x,..,x)$ associée à une application multilinéaire symétrique $F(x_1,..,x_n)$.
On va formuler ce dernier point d'une manière plus précise.

Applications multilinéaires et polynomes

Soient X et Y deux espaces de Banach complexes; notons $\mathcal{L}(X^n;Y)$ l'espace de Banach $\mathcal{L}(X,..,X;Y)$. Le sous espace vectoriel de $\mathcal{L}(X^n;Y)$, composé des applications n-linéaires symétriques continues de $X\times..\times X$ dans Y, on le notera $\mathcal{L}_s(X^n;Y)$, est un sous espace de Banach de $\mathcal{L}(X^n;Y)$.

L'espace $\mathcal{P}(X^n;Y)$ des polynomes homogènes de degré n continus sur X et à valeurs dans Y, est aussi un espace de Banach muni de la norme

$$\|P\| = \sup\{\|P(x)\| \ / \ x \in X, \ \|x\| \leq 1\} \quad \text{pour } P \in \mathcal{P}(X^n;Y)$$

L'application naturelle :

$$A \in \mathcal{L}_s(X^n;Y) \to P_A \in \mathcal{P}(X^n;Y) \quad \text{où } P_A(x)=A(x,..,x)$$

établit un isomorphisme algébrique et topologique entre les deux espaces.

Remarquons qu'on peut toujours associer à une application multilinéaire continu une application multilinéaire symétrique; pour cela on associe à $A\in\mathcal{L}(X^n;Y)$ l'application $A_s\in\mathcal{L}_s(X^n;Y)$ définie par :

$$A_s(x_1,..,x_n) = \frac{1}{n!} \sum_{\sigma\in\mathcal{O}_n} A(x_{\sigma(1)},..,x_{\sigma(n)})$$

$x_i \in X \ i=1,..,n$
$\mathcal{O}_n$ = groupe symétrique degré n

L'application $A \to A_s$ de $\mathcal{L}(X^n;Y)$ dans $\mathcal{L}_s(X^n;Y)$ est linéaire et continue.

Exemple : Les opérateurs H-S

Soient $X=Y=L^2(T)$, T un intervalle réèl, $k \in L^2(T^{n+1})$ et $x_1(s_1)$, $..,x_n(s_n)$ appartenant à X. L'application :

$$K(x_1,..,x_n) = \int_{T^n} k(t,s_1,..,s_n)x_1(s_1)..x_n(s_n)ds_1..ds_n$$

définit un opérateur n-linéaire, continu de X^n dans Y, symétrique si le noyau k l'est; à K il correspond le polynome :

$$P(x) = K(x,..,x) = \int_{T^n} k(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

Si K n'est pas symétrique, on peut d'après ce qui précède lui associer une application symétrique K_s qui donnera le polynome :

$$P(x) = \int_{T^n} \sum_{\pi \in \mathfrak{S}_n} \frac{k(t,\pi s)}{n!} x(s_1)..x(s_n)ds_1..ds_n$$

où $s=(s_1,..,s_n)$ et π=permutation des composantes de s.
On aura toujours :

$$P(x) = K_s(x,..,x) = K(x,..,x)$$

car les intégrales obtenues par permutation sont toutes égales entre elles.

3. Opérateurs polynomes H-S

Soit T un intervalle réèl; considérons les opérateurs H-S A de $L^2(T^n)$ dans $L^2(T)$; on a vu qu'on pouvait les représenter par :

$$(Ax)(t) = \int_{T^n} k(t,s)x(s)ds \qquad t \in T \quad s=(s_1,..,s_n) \in T^n$$

Les combinaisons linéaires de tels opérateurs avec $x(s)=x(s_1)..x(s$ seront dits opérateurs polynomes H-S.
Un polynome H-S de degré N s'écrira :

$$(Hx)(t) = \sum_{n=0}^{N} \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

P A R T I E 1

APPROXIMATION D'OPERATEURS NON LINEAIRES

chap.1 : Approximation d'opérateurs non linéaires continus

- opérateur défini dans un L^2 : approximation par un polynome H-S
- opérateur défini dans un Hilbert séparable : approximation par un opérateur polynome.

chap.2 : Opérateurs non linéaires analytiques

- cas général : représentation par une série de polynomes
- opérateurs définis dans un L^2 : représentation par une série de polynomes H-S.

CHAPITRE 1

APPROXIMATION D'OPERATEURS NON LINEAIRES CONTINUS

1.1. Opérateurs définis dans un L^2

Un opérateur polynome H-S de degré N est un opérateur H_N défini par (chapitre 0) :

$$(H_N x)(t) = \sum_{n=0}^{N} \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n \qquad (1)$$

Presque tout ce chapitre sera consacré à la démonstration de la densité des opérateurs polynomes H-S dans l'ensemble des applications non linéaires définies et continues sur un compact de $L^2(T)$ (T intervalle réèl) et à valeurs dans $L^2(T)$.

On montrera que si A est un opérateur non linéaire dans $L^2(T)$, continu sur un compact $X \subset L^2(T)$, on peut l'approximer par un polynome H-S H_N ayant la forme (1) :

$\forall \varepsilon > 0 \; \exists H_N$ (le degré N et les noyaux k_n dépendent de ε) tel que :

$$\| A - H_N \|_{C[X,L^2(T)]} = \sup_{x \in X} \| Ax - H_N x \|_{L^2(T)} \leq \varepsilon$$

H_N peut être choisi de telle sorte que ponctuellement on ait une inégalité analogue :

Si t est un réèl du domaine de définition de la fonction (Ax), on a presque partout (l'ensemble des t où elle n'est pas vérifiée est de mesure nulle) :

$$|(Ax)(t) - (H_N x)(t)| \leq \varepsilon$$

(en effet, de la suite (H_N) qui tend vers A en norme, on peut extraire une suite qui converge presque partout vers A; les éléments (H_n) de cette suite extraite satisfont les deux inégalités : presque partout et en norme. H_N peut donc être pris parmi les éléments de cette sous suite).

Soient T un intervalle réèl, $t \in T$, X un compact de $L^2(T)$ et $C[X,L^2(T)]$ l'e.v.n. des fonctions continues de X dans $L^2(T)$ (muni de la norme du sup).

Notons $\mathcal{H}_n$ l'ensemble des polynomes H-S de degré au plus égal à n, définis sur X :

$$H \in \mathcal{H}_n \quad \leftrightarrow \quad (Hx)(t) = \sum_{j=0}^{n} \int_{T^j} k_j(t,s_1,..,s_j)x(s_1)..x(s_j)ds_1..ds_j \qquad x \in X$$

On supposera les k_j symétriques par rapport à $s_1,..,s_n$.
Par définition même des opérateurs H-S : $\mathcal{H}_n \subset C[X,L^2(T)]$

Théorème de représentation :

Soit $\mathcal{K} \subset C[X,L^2(T)]$, avec $\mathcal{K}$ relativement compacte ($\bar{\mathcal{K}}$ compacte); alors, $\forall \varepsilon>0 \ \exists N=N(\varepsilon)$ tel que :
$\forall F \in \bar{\mathcal{K}}$, il existe un polynome H-S : $H \in \mathcal{H}_N$ satisfaisant la relation :

$$\|F - H\|_{C[X,L^2(T)]} = \sup_{x \in X} \|Fx - Hx\|_{L^2(T)} \leqslant \varepsilon$$

Presque tout ce chapitre sera consacré à la démonstration de ce théorème. Remarquons qu'il suffit de l'établir pour les éléments de $\mathcal{C}$ car d'une part, $\mathcal{C}$ et $\bar{\mathcal{C}}$ ayant même topologie, $\forall F \in \bar{\mathcal{C}}$, $\exists F_1 \in \mathcal{C}$ suffisament proche de F (les éléments de $\mathcal{C}$ sont partout denses dans $\bar{\mathcal{C}}$); on prendra F_1 tel que : $\|F - F_1\| < \frac{\varepsilon}{2}$.

D'autre part, si le résultat est établi pour les éléments de $\mathcal{C}$ $\forall F_1 \in \mathcal{C}$ $\exists H$ polynome HS tel que $\|F_1 - H\| < \frac{\varepsilon}{2}$

Donc $\forall F \in \bar{\mathcal{C}}$, $\exists H \in \mathcal{H}_N$ tel que :

$$\|F - H\| \leq \|F - F_1\| + \|F_1 - H\| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

Donc, si le résultat est établi pour $\mathcal{C}$, il sera vrai pour $\bar{\mathcal{C}}$.

On va d'abord démontrer le théorème pour les fonctionnelles continues. Appellons $\mathcal{H}$ la classe de toutes les fonctionnelles de type HS :

$$z = K(x) = \sum_{j=0}^{n} \int_{T^j} k_j(s_1,..,s_j)x(s_1)..x(s_j)ds_1..ds_j \qquad (2)$$

$$k_j \in L^2(T^j) \quad z \in R$$

Lemme :

Soit F une fonctionnelle sur X compact ($F : X \subset L^2(T) \to R$); il existe alors une suite d'éléments de $\mathcal{H}$, c'est à dire des fonctionnelles H-S du type (2), qui converge uniformément vers F.

En effet :

- $\mathcal{H}$ est une algèbre.
- $\mathcal{H}$ sépare les points de X car si $x_1 \neq x_2$, il existe un noyau k_1 tel que :

$$\int_T k_1(s)x_1(s)ds \neq \int_T k_1(s)x_2(s)ds$$

- Toute fonctionnelle appartenant à $\mathcal{H}$ est évidemment continue.

D'après le théorème de Stone-Wierstrass, $\forall \varepsilon>0$, il existe un entier $N(\varepsilon)=N$ tel que pour toute fonctionnelle F continue sur X, on ait :

$$|F(x) - K(x)| = \left|F(x) - \sum_{j=0}^{N} \int_{T^j} k_j(s_1,..,s_j)x(s_1)..x(s_j)ds_1..ds_j\right| < \varepsilon$$

pour tout $x \in X$.

Preuve du théorème :

Soit $F \in \mathcal{F} \subset C[X,L^2(T)]$ avec X et $\bar{\mathcal{F}}$ compacts et $X \subset L^2(T)$.
$F \in \mathcal{F}$, l'ensemble F(X) est précompact.†

Posons : $$\mathcal{F}(X) = \bigcup_{F \in \mathcal{F}} F(X) \subset L^2(T)$$

$\mathcal{F}$ précompact $\rightarrow \mathcal{F}(X)$ l'est aussi.

- $\mathcal{F}(X)$ est un sous ensemble d'un Hilbert séparable $L^2(T)$; alors, quel que soit le choix d'un système orthonormé complet $\{\psi_k\}$ dans $L^2(T)$, pour $\varepsilon>0$ donné, $\exists K=K(\varepsilon)$ tel que pour tout $y \in \mathcal{F}(X)$ on ait :

$$\left|\sum_{K+1}^{\infty} (y,\psi_k)\psi_k\right|^2 = \sum_{K+1}^{\infty} |(y,\psi_k)|^2 < \varepsilon^2 \qquad (3)$$

- Fixons $K=K(\varepsilon)$ et notons $\mathcal{F}(X)_K$ l'image de $\mathcal{F}(X)$ par la projection π :

$$\pi : y \rightarrow \sum_{k=1}^{K} (y,\psi_k)\psi_k$$

$\mathcal{F}(X)_K$ est alors précompact.

Remarquons que si F_1 et $F_2 \in \mathcal{F}$ et satisfont $\|F_1 - F_2\| > 3\varepsilon$, on a alors pour tout $x \in X$:

$$\|\pi F_1 x - \pi F_2 x\| > \varepsilon \qquad (4)$$

† X espace métrique, une partie M de X est précompacte si $\forall \varepsilon>0$, M peut être recouverte par un nombre fini de boules de rayon $<\varepsilon$ et dont les centres $\in M$.
X complet, M précompacte $\leftrightarrow$ $\bar{M}$ compacte.

Démontrons (4) :

$$3\varepsilon < \|F_1 - F_2\| = \|F_1 - F_2 + \pi F_1 x - \pi F_1 x + \pi F_2 x - \pi F_2 x\|$$

$$\leqslant \|F_1 - \pi F_1 x\| + \|F_2 - \pi F_2 x\| + \|\pi F_2 x - \pi F_1 x\|$$

$$3\varepsilon - \|F_1 - \pi F_1 x\| - \|F_2 - \pi F_2 x\| < \|\pi F_2 x - \pi F_1 x\|$$

D'après (3) et la définition de $F_1 x$:

$$\|F_1 - \pi F_1 x\| = \| \sum_1^{\infty} - \sum_1^{K} \| = \| \sum_{K+1}^{\infty} \| < \varepsilon$$

De même $\|F_2 - \pi F_2 x\| < \varepsilon$

D'où : $3\varepsilon - \varepsilon - \varepsilon = \varepsilon \leqslant \|\pi F_2 x - \pi F_1 x\|$

- Revenons à l'ensemble $\mathcal{K}$

$\mathcal{K}$ précompact $\longrightarrow$ il peut être recouvert par un nombre fini L de boules de diamètre ε ; mais ces boules sont composées aussi d'éléments de $\mathcal{K}(X)$ (algébriquement, l'ensemble des trajectoires peut être identifié avec l'ensemble des points de ces trajectoires).

Proposition :

Considérons $F_1, .., F_L$ éléments arbitraires de $C[X, L^2(T)]$, ces éléments étant tels que F_i appartienne à la $i^{ème}$ boule$(i=1,..,L)$.
Alors, pour un certain entier N, on peut trouver $H_1, .., H_L$ tels que : $H_i \in \mathcal{H}_N$

$$\|F_i - H_i\| < 3\varepsilon \qquad i=1,..,L \qquad (5)$$

Cette proposition , qu'on démontre plus loin, achèvera la démonstration du théorème :

En effet, $F \in \mathcal{K}$, $\mathcal{K}$ étant recouvert par L boules, F appartiendra nécessairement à l'une de ces boules, la $i^{ème}$ par exemple. F se situant

dans la même boule que F_i, on aura $\|F-F_i\| < \varepsilon$, et le H_i (déterminé par (5)) qui approximera F_i approximera aussi F puisque :

$$\|F - H_i\| < \|F - F_i\| + \|F_i - H_i\| < 4\varepsilon$$
$$< \varepsilon \qquad < 3\varepsilon$$

Pour démontrer la proposition, on va utiliser le théorème déjà établi pour les fonctionnelles de $C[X,R]$.
Remarquons que chaque $(\psi_i, F_j x)$ est une fonctionnelle sur X àvaleurs réèlles et continue, et cela pour i=1,..,K j=1,..,L

On a déjà vu que la classe $\mathcal{H}$ des fonctionnelles HS du type (2) était partout dense dans $C[X,R]$.
Donc, pour toute fonctionnelle $(\psi_i, F_j x)$ et $\forall \varepsilon' > 0$, il existe un entier N_{ij} et un ensemble de noyaux $k_n^{(i,j)}$ $(n=1,..,N_{ij})$ définissant des fonctionnelles H-S K^{ij} telles que :

$$|(\psi_i, F_j x) - K^{(i,j)}(x)| = \tag{6}$$
$$= \left|(\psi_i, F_j x) - \sum_{n=0}^{N_{ij}} \int_{T^n} k_n^{(i,j)}(s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n\right| < \varepsilon'$$

et cela $x \in X$; prenons $\varepsilon' = \frac{\varepsilon}{K}$.

- Rappelons que dans la proposition on cherche N entier de façon à avoir (5).
Plus précisément, $F_1,..,F_L$ étant des éléments de $C[X,L^2(T)]$ tels que F_i appartienne à la $i^{ème}$ boule, on cherche N et $H_1,..,H_L \in \mathcal{H}_N$ tels que :

$$\|F_i - H_i\| < 3\varepsilon \qquad i=1,...,L$$

Les H_i ainsi que N seront déterminés respectivement à partir de $K^{(i,j)}$ et N_{ij} obtenus dans (6), et cela par la construction suivante :

Soient $k_n^{(i,j)}$ les noyaux obtenus pour les fonctionnelles dans (6), et $K^{(i,j)}(x)$ les fonctionnelles H-S associées.
Si on prend un système orthonormé complet dans $L^2(T)$, $k_n^{(i,j)}$ peut se représenter par la série de Fourier :

$$k_n^{(i,j)} = \Sigma\, a^{(i,j)}_{i_1 \ldots i_n} \phi^{(i_1)}(s_1) \ldots \phi^{(i_n)}(s_n) \qquad (7)$$

La fonctionnelle H-S $K^{(i,j)}$ associée aux noyaux (7) $(n=0,1,..,N_{ij})$ approxime la fonctionnelle $(\psi_i, F_j x)$.

Prenons $N_j = \max N_{ij}$ $i=1,..,K$ et considérons cette fois-ci l'opérateur H-S $H_j \in \mathcal{H}_{N_j}$ défini par les noyaux :

$$h_n^{(j)} = \sum_{k,i_1,..,i_n} b^{(j)}_{k,i_1,..,i_n} \psi_k(t)\phi_{i_1}(s_1)..\phi_{i_n}(s_n)$$

$$n=0,1,..,N_j$$

- Nous avons $(\psi_i, H_j x) = K^{(i,j)}(x)$ pour chaque i si :

$$b^{(j)}_{i,i_1,..,i_n} = a^{(i,j)}_{i_1,..,i_n}$$

pour i=1,..,K (les valeurs de b pour les autres i n'interviennent pas) et pour tous les $i_1,\ldots,i_n$.

H_j étant construit de la sorte, (en égalant les coéfficients a et b, on peut remplacer $K^{(i,j)}(x)$ par $(\psi_i, H_j x)$ dans la relation (6); on obtient alors :

$$\left|(\psi_i, F_j x) - (\psi_i, H_j x)\right| < \frac{\varepsilon}{K} \qquad \forall\, x \in X \quad \text{et } i=1,..,K$$

En appliquant cette dernière relation, on obtient :

$$\left|(\psi_1,F_jx) +\ldots+ (\psi_K,F_jx) - (\psi_1,H_jx) -\ldots- (\psi_K,H_jx)\right|$$

$$\leq \left|(\psi_1,F_jx) - (\psi_1,H_jx)\right| +\ldots+ \left|(\psi_K,F_jx) - (\psi_K,H_jx)\right| < \frac{\varepsilon}{K} +\ldots+ \frac{\varepsilon}{K} = \varepsilon$$

Nous avons donc pour tous les $x \in X$:

$$\left|\pi F_jx - \pi H_jx\right| < \varepsilon \qquad (8)$$

Or, on a vu en (4) que $\|F_1-F_2\| > 3\varepsilon \rightarrow \|\pi F_1x-\pi F_2x\| > \varepsilon$, ou ce qui est équivalent

$$\|F_1x-\pi F_2x\| < \varepsilon \rightarrow \|F_1-F_2\| < 3\varepsilon$$

En appliquant cette dernière relation à (8), on a alors :

$$\|F_j - H_j\| < 3\varepsilon$$

On peut donc prendre $N = \max N_j$ où $j=1,\ldots,L$.

Ceci achève la démonstration du théorème.

1.2. Opérateurs définis dans un Hilbert séparable

On va énoncer un théorème de densité analogue à celui qu'on vient de démontrer mais avec un Hilbert séparable à la place d'un L^2, et un opérateur polynome au lieu d'un opérateur H-S.

Soient X et Y deux e.v.n., K une partie compacte de X, et C(K,Y) l'e.v.n. des applications continues de K dans Y.
Le théorème de Weierstrass établit que si X=Y=R, la famille des polynomes est dense dans C(K,R).
Ce théorème a une généralisation si X=Y=H, où H est un Hilbert séparable:

Théorème : [4]

La famille des opérateurs polynomes dans H définis et continus dans un compact K de H est dense dans C(K,H).

Donc, si A est une application continue de K dans H :
$\forall \varepsilon > 0$ il existe un opérateur polynome $P_N^{(\varepsilon)}$ tel que :

$$\| A - P_N^{(\varepsilon)} \| = \sup_{x \in K} \| Ax - P_N^{(\varepsilon)} x \| \leqslant \varepsilon$$

Ceci est cité à titre de généralisation ; toutefois il ne présente pas de possibilités d'exploitation numérique car $P_N^{(\varepsilon)}$ a une forme très générale:

$$P_N(x) = L_0 + L_1 x + \ldots + L_N x^N$$

(L_0= application constante; L_n= application n-linéaire de $H^n \to H$;
$L_n x^n = L_n(x,..,x)$).

1.2. Opérateurs définis dans un Hilbert séparable

On va énoncer un théorème de densité analogue à celui qu'on vient de démontrer mais avec un Hilbert séparable à la place d'un L^2, et un opérateur polynome au lieu d'un opérateur H-S.

Soient X et Y deux e.v.n., K une partie compacte de X, et C(K,Y) l'e.v.n. des applications continues de K dans Y.
Le théorème de Weierstrass établit que si X=Y=R, la famille des polynomes est dense dans C(K,R).
Ce théorème a une généralisation si X=Y=H, où H est un Hilbert séparable:

Théorème :[4]

La famille des opérateurs polynomes dans H définis et continus dans un compact K de H est dense dans C(K,H).

Donc, si A est une application continue de K dans H :
$\forall \varepsilon > 0$ il existe un opérateur polynome $P_N^{(\varepsilon)}$ tel que :

$$\| A - P_N^{(\varepsilon)} \| = \sup_{x \in K} \| Ax - P_N^{(\varepsilon)} x \| \leqslant \varepsilon$$

Ceci est cité à titre de généralisation ; toutefois il ne présente pas de possibilités d'exploitation numérique car $P_N^{(\varepsilon)}$ a une forme trés générale:

$$P_N(x) = L_0 + L_1 x + \ldots + L_N x^N$$

(L_0= application constante; L_n= application n-linéaire de $H^n \to H$;
$L_n x^n = L_n(x,..,x)$).

CHAPITRE 2

APPROXIMATION D'OPERATEURS ANALYTIQUES DEFINIS DANS UN L^2

On a vu qu'un opérateur non linéaire A défini et continu dans un L^2 (resp. un Hilbert séparable) pouvait être approché par un opérateur P_N^ε du type H-S (resp. un opérateur polynome). Le changement de ε fournit évidemment un autre polynome $Q_{N'}^{\varepsilon'}$, en général totalement différent de P_N^ε.

On va étudier le cas où en plus de la continuité de A, on suppose son analycité. La représentation polynomiale obtenue devient intéressante dans ce cas car si on change ε, le nouveau polynome $P_{N'}^{\varepsilon'}$ approximant A diffère de P_N^ε seulement par le nombre de termes, les polynomes homogènes de même degré de P_N^ε et $P_{N'}^{\varepsilon'}$ sont identiques.

Dans le cas général, un opérateur non linéaire et analytique A défini dans un e.v.n. admet une représentation par une série de polynomes homogènes, la série de Taylor (les applications dérivées sont multilinéaires et deviennent des polynomes homogènes lorsqu'elles sont appliquées au même vecteur $(x,..,x)\equiv x^n$).

L'objet de ce chapitre est d'étudier le cas où A est analytique défini dans $L^2(T)$, et d'en chercher une représentation par une série de polynomes H-S :

$$(Ax)(t) = \sum_{n=0}^{\infty} \int_{T^n} k_n(t,s_1,.,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

Aprés avoir établi une généralisation du théorème de Riesz aux fonctionnelles multilinéaires continues dans $L^2(T)$ (représentation intégrale à noyau dans $L^2(T^n)$), et en s'appuyant sur les résultats du cas général, on

prouvera d'abord la représentation pour les fonctionnelles analytiques ; le passage à l'opérateur A se fera en considérant celui-ci comme une famille de fonctionnelles A_t définies par :

$$A_t(x) = (Ax)(t)$$

On examinera la validité de cette écriture et le problème de convergence qu'elle soulève lors du passage de A_t à A.

2.1. Analycité dans un e.v.n.

Soient X et Y deux e.v.n., O un ouvert de X et A: O → Y.

Rappelons que la dérivée de A au point a de O est l'application linéaire continue L $\in \mathcal{L}C(X,Y)$ telle que :

$$A(a+h) - A(a) = L(h) + \|h\|\varepsilon(a,h) \qquad a \text{ et } h \in X$$

$$\varepsilon: X\times X \to Y,\ \varepsilon \to 0 \text{ lorsque } h \to 0$$

La dérivée seconde $A''(a)=(A'(a))' \in \mathcal{L}(X;\mathcal{L}(X,Y)) \sim \mathcal{L}_2(X,X;Y)$ est une application bilinéaire continue de X×X dans Y, et d'une manière générale $A^{(n)}(a)$ est une application n-linéaire de X^n dans Y.

Si A possède des dérivées jusqu'à l'ordre n-1 dans O, une dérivée d'ordre n en a, on a alors en notant $A^{(n)}(a)(h,..,h)\equiv A^{(n)}(a)h^n$, le développement de Taylor :

$$A(a+h)= A(a)+ A'(a)h +...+ \frac{A^{(n)}}{n!}(a)h^n +\alpha\|h\|^n \qquad \alpha \to 0 \text{ avec } h$$

et l'analycité se définit par analogie avec le cas des fonctions numériques:

A est analytique dans un domaine $\mathcal{D}$ si les dérivées de tous les ordres existent dans $\mathcal{D}$ et si la série du développement de Taylor converge normalement et uniformément dans $\mathcal{D}$.

développement polynomial d'un opérateur

L'application A analytique admet un développement en une série de dérivées $A^{(n)}$ qui sont des applications multilinéaires de X^n dans Y, et fournissent des polynomes homogènes lorsqu'elles sont appliquées au même vecteur h. [3]

A admet donc le développement :

$$A(x+h)= \sum_{n=0}^{\infty} a_n P_n(x,h)$$

avec $a_n= \frac{1}{n!}$ et $P_n(x,h)$ polynome en h de degré n; le développement converge normalement et uniformément sur tout compact contenu dans le domaine d'analycité.

Représentation d'un opérateur par une famille de fonctionnelles

Soit A un opérateur analytique défini dans $L^2(T)$; pour les besoins de l'étude, on sera amené à le considérer comme une famille de fonctionnelles A_t définies par :

$$A_t(x) = (Ax)(t) \qquad x \in L^2(T) \ , \ t \in T \subset R \qquad (1)$$

La représentation (1) de A pose un problème :

En effet, (Ax) est un élément de L^2 et représente ainsi une classe de fonctions égales presque partout; on ne peut donc lui attribuer une valeur précise en un point fixé t; de ce fait $A_t(x)$ ne peut être définie d'une manière précise .

Cette imprécision disparait si (Ax) est une fonction continue, et la représentation obtenue pour de telles fonctions demeurera valable pour les opérateurs dans $L^2(T)$. Précisons ces deux points :

- Pour que la représentation (1) ait un sens, il suffit que la fonction (Ax) soit continue; A devra donc être défini dans le sous ensemble :

$$E = \{ x \in L^2(T) \ / \ Ax \text{ continue} \}$$

- Il faudrait ensuite pouvoir conclure que les opérateurs A et $\tilde{A} = \{ A_t \}_{t \in T}$ coïncidant dans E, coincident dans $L^2(T)$.
(l'égalité dans L^2 étant presque partout) Il suffit que A et $\tilde{A}$ coincident dans une partie dense de $L^2(T)$. Le domaine de définition E de A sera alors pris de telle sorte que : $\bar{E} = L^2(T)$.

A devra donc être défini dans le sous ensemble E de $L^2(T)$ défini par :

$$E = \{ x \ / \ x \in L^2(T), Ax \text{ continue}, \bar{E} = L^2(T) \}$$

On pourra prendre pour E l'ensemble des fonctions dérivables ou des fonctions indéfiniment dérivables (ensembles denses dans $L^2(T)$), et "raisonnablement", on peut penser que parmi les opérateurs analytiques, les opérateurs conservant au moins la continuité des éléments de E forment une classe assez large.

2.2. Approximation d'opérateurs non linéaires analytiques dans $L^2(T)$

Soit A un opérateur non linéaire et analytique dans un domaine de $L^2(T)$; on va montrer que moyennant certaines conditions, A admet une représentation par une série d'opérateurs H-S.

Le résultat sera d'abord établi pour une fonctionnelle analytique, le passage à l'opérateur A se fera en considérant A comme une famille de fonctionnelles A_t.

2.2.1. Représentation des fonctionnelles analytiques

Soit A une fonctionnelle analytique : $L^2(T) \to R$; on sait qu'elle admet un développement en série de Taylor dont le terme général :

$$\frac{1}{n!}A^{(n)}(.)x^n = \left[\frac{1}{n!}A^{(n)}(.)\right](x,..,x)$$

est une fonctionnelle multilinéaire continue.

Pour obtenir la représentation de A, on va montrer qu'une fonctionnelle multilinéaire continue admet une représentation intégrale du type HS :

$$F(x_1,..,x_n) = \int_{T^n} k(s_1,..,s_n)x_1(s_1)..x_n(s_n)ds_1..ds_n$$

$$k \in L^2(T^n) \ , \ x_j \in L^2(T)$$

Ce résultat appliqué à la fonctionnelle $F= \frac{1}{n!}A^{(n)}(.)$ au point $(x,..,x)=x^n$ donne une fonctionnelle polynome H-S :

$$\int_{T^n} k_n(s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

d'où en remplaçant dans le developpement de Taylor de A :

$$A(x) = \sum_{n=0}^{\infty} \int_{T^n} k_n(s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

Représentation d'une fonctionnelle multilinéaire continue

Soit H(T) le sous ensemble de $L^2(T)$ composé des fonctions f dérivables et telles que $|f|^2$ et $|f'|^2 \in L^2(T)$. H(T) est muni de la norme de $L^2(T)$.

Théorème :

Soit $F(x_1, ..,x_n)$ une fonctionnelle multilinéaire continue sur H(T). Si :

$$|F(x_1,..,x_j',..,x_k',..,x_n)| \leq M\|x_1\|..\|x_n\| \qquad j=1,..,n \; k=1,..,n$$

pour toutes les fonctions $x_j \in H(T)$,

il existe alors une fonction $k \in L^2(T^n)$ telle que :

$$F(x_1,...,x_n) = \int_{T^n} k(s_1,..,s_n)x_1(s_1)..x_n(s_n)ds_1..ds_n$$

Ce théorème généralise le théorème de Riesz aux fonctionnelles multilinéaires continues.
La condition imposée par l'inégalité diffère de la continuité par l'apparition de deux dérivées, et est assez faible (elle est satisfaite par exemple par toutes les fonctions polynomes, sinusoidales, exponentielles).

Lemme : (Guelfand) [1]

Soit $B(x_1,x_2)$ une fonctionnelle bilinéaire continue sur $H(T)$ et telle que :

$$|B(x_1',x_2')| \leqslant M\|x_1\|\|x_2\|$$
$$|B(x_1,x_2')| \leqslant M\|x_1\|\|x_2\|$$
$$|B(x_1',x_2)| \leqslant M\|x_1\|\|x_2\|$$

La fonctionnelle B a alors la forme :

$$B(x_1,x_2) = \int_{T2} k(t_1,t_2)x_1(t_1)x_2(t_2)dt_1dt_2 \qquad \text{avec } k \in L^2(T^2)$$

Généralisation du lemme :

Le lemme reste valable pour les fonctions de plusieurs variables, c'est à dire pour $x_j \in H(T^n)$, mais les inégalités du lemme devront être remplacées par les suivantes :

$$|B(x_1^{(j)},x_2^{(k)})| \leqslant M\|x_1\|\|x_2\| \qquad 0 \leqslant j \leqslant n \qquad 0 \leqslant k \leqslant n$$

$x_1^{(j)}$ étant la dérivée simple par rapport à la $i^{ème}$ variable.

Dans ces conditions, toute fonctionnelle bilinéaire continue satisfaisant les inégalités admet une représentation intégrale avec un noyau appartenant à $L^2(T^{2n})$.

Preuve du théorème :

On le démontre par récurrence.

n=1 : on a le théorème de Riesz.

Supposons le résultat vrai pour p, c'est à dire que toute fonctionnelle p-linéaire continue sur H(T) satisfaisant les inégalités s'écrit :

$$F(x_1,..,x_p) = \int_{Tp} k(t_1,..,t_p)x_1(t_1)..x_p(t_p)dt_1..dt_p \qquad (a)$$

et montrons le pour p+1, c'est à dire que toute fonctionnelle p+1-linéaire sur H(T) s'écrit :

$$F(x_1,..,x_{p+1}) = \int_{T^{p+1}} k(t_1,..,t_{p+1})x_1(t_1)..x_{p+1}(t_{p+1})dt_1..dt_{p+1} \qquad (b)$$

Posons

$$\tilde{t}_p = (t_1, \ldots, t_p)$$

$$d\tilde{t}_p = dt_1 \ldots dt_p$$

$$\tilde{x}(\tilde{t}) = x_1(t_1) \ldots x_p(t_p)$$

L'expression (a) s'écrit avec ces notations :

$$\int_{\tilde{T}} K(\tilde{t}_p)\, \tilde{x}(\tilde{t})\, d\tilde{t}_p \qquad \text{avec} \quad \tilde{x}(\tilde{t}) \in H^1(T^p) \qquad (a'$$

Le problème est alors de montrer, avec les nouvelles notations, que toute forme (p+1) - linéaire continue sur $H(T^{p+1})$ peut s'écrire

$$\int_{\tilde{T}\times T} K(\tilde{t}_p, t_{p+1})\, \tilde{x}(\tilde{t})\, x_{p+1}(t_{p+1})\, d\tilde{t}_p\, dt_{p+1} \qquad (b')$$

L'expression (b') ressemble à la représentation d'une fonctionnelle bilinéaire, à cela près, que $\tilde{x}$ et x_{p+1} n'appartiennent pas au même espace. Pour essayer de se ramener, au cas d'une fonctionnelle bilinéaire, il faudrait que $\tilde{x}(\tilde{t})$ et $x_{p+1}(t_{p+1})$ puissent être considérés comme appartenant au même espace H . Cela est possible car :

- $\tilde{x}(\tilde{t}) \in H(T^p)$

  en effet $\tilde{x}(\tilde{t}) = x_1(t_1) \ldots x_p(t_p)$

  $\tilde{x}$ est de carré de module sommable car

$$\int_{T^p} |\tilde{x}(\tilde{t})|^2 d\tilde{t} = \int_{T^p} |x_1(t_1)|^2 \ldots |x_p(t_p)|^2 dt_1 \ldots dt_p$$

$$= \int_T |x_1(t_1)|^2 dt_1 \int_T |x_2(t_2)|^2 dt_2 \ldots \int_T |x_p(t_p)|^2 dt_p$$

Comme $x_j \in L^2(T) \Rightarrow \tilde{x}(t) \in L^2(T^p)$

En appliquant la règle de dérivation d'un produit et le même raisonnement on montre que $\tilde{x}(\tilde{t}) \in H(T^p)$

- $x_{p+1}(t_{p+1}) \in H(T) \Longrightarrow$ on peut considérer que $x_{p+1}(t_{p+1}) \in H(T^p)$

cela est vrai en identifiant les deux fonctions :

$$t_{p+1} \longrightarrow x_{p+1}(t_{p+1})$$

et

$$(t_2,\ldots,t_{p+1}) \longrightarrow \left(0,\ldots, 0, x_{p+1}(t_{p+1})\right)$$

ces deux fonctions peuvent en effet être identifiées puisqu'elles prennent des valeurs que l'on peut considérer comme identiques, quel que soit le vecteur $(t_2,\ldots,t_{p+1}) \in T^p$.

Elles ont même norme aussi, donc, si $x'_{p+1}(t_{p+1}) \in L^2(T)$, la deuxième fonction appartient à $L^2(T^p)$.

De même

$$x'_{p+1}(t_{p+1}) \sim (0, \ldots,, 0, x'_{p+1}(t_{p+1}))$$

et si $x'_{p+1} \in L^2(T)$, alors la dérivée de la deuxième fonction appartient à $L^2(T^p)$.

> On peut donc identifier $x_{p+1}(t_{p+1})$ élément de H (T) à un élément de $H(T^p)$.

Revenons à la récurrence ; montrer que toute forme (p+1) - linéaire continue sur H (T) admet une représentation intégrale revient à montrer toute fonctionnelle bilinéaire continue sur $H(T^p)$ admet la représentation intégrale (b'). Cela découle de la généralisation du lemme.

Donc le théorème fondamental est démontré.

2.2.2. Représentation des opérateurs analytiques

A est un opérateur analytique dans $L^2(T)$; le résultat obtenu pour la représentation d'une fonctionnelle analytique s'énonce ainsi :

Pour t fixé, la série numérique :

$$\sum_{n=0}^{\infty} \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

converge vers le nombre $A_t(x)$.

Si t devient variable, la représentation obtenue pour la fonctionnelle A_t peut se généraliser à l'opérateur A (qu'on assimilera à une famille de fonctionnelles A_t); il suffit pour cela, de montrer que la série des fonctions (de $L^2(T)$) :

$$\sum_{n=0}^{\infty} u_n(t) = \sum_{n=0}^{\infty} \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

converge en norme L^2 vers la fonction $(Ax)(t)$.
Nous allons établir cette convergence en imposant de nouvelles conditions à A.

$$\text{Posons : } S_n = \sum_{j=0}^{n} u_j(t) \quad \text{et} \quad S = \sum_{j=0}^{\infty} u_j(t)$$

Rappelons que :

$$\sum_{j=0}^{\infty} \|u_j\|_{L^2} \leq M < +\infty \longrightarrow \|S_n - S\|_{L^2} \to 0$$

(en effet, si $\sum_{0}^{\infty} \|u_j\|_{L^2} < +\infty$, alors la série $\sum_{0}^{\infty} u_j(t)$ converge presque partout, définit une fonction qui appartient à L^2 (fonction qui dans ce cas est S puisqu'il y a convergence partout), et la convergence a lieu dans L^2) [5]

On connait la fonction vers laquelle $\sum_{j=0}^{\infty} u_j(t)$ converge presque partout : c'est la fonction $S(t)$.

Il y a même convergence partout de $S_n(t)$ vers $S(t)$:

en effet, pour tout t fixé, on a une série numérique :

$$\sum_{j=0}^{\infty} \int_{T^j} k_j(t,s_1,..,s_j)x(s_1)..x(s_j)ds_1..ds_j$$

qui, on l'a déjà vu, converge vers le nombre $A_t(x)$.

Or, par définition

$$A_t(x) = (Ax)(t)$$

et ainsi, la série

$$\sum_{0}^{\infty} u_j(t)$$

converge partout vers : $(Ax)(t)$.

Montrons que la série de fonctions $\sum_{0}^{\infty} u_n(t)$ converge dans L^2; il suffit de montrer que $\sum_{0}^{\infty} \|u_n\|$ converge; cela est possible en imposant de nouvelles conditions à l'opérateur A.

Moyennant ces conditions, la représentation de A_t se généralise à A, et on pourra écrire :

$$(Ax)(t) = \sum_{n=0}^{\infty} \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

conditions de convergence

On va déterminer les conditions pour que la série $\sum_{n=0}^{\infty} u_n(t)$ converge dans L^2.

En fait, A étant analytique, $u_n(t) = \frac{1}{n!} \int k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$

($\frac{1}{n!}$ est le coéfficient de la dérivée d'ordre n, $A^{(n)}$, dans le développement de Taylor; jusqu'alors, il n'était pas utilisé et était intégré dans le noyau k_n)

Etudions la convergence de $\sum_{n=0}^{\infty} \|u_n\|$:

Posons :

$$s=(s_1,..,s_n) \qquad ds = ds_1..ds_n \qquad x(s)= x(s_1)..x(s_n)$$

$$\|u_n\| = \frac{1}{n!}\left[\int_T \left|\int_{T^n} k_n(t,s)x(s)ds\right|^2 dt\right]^{\frac{1}{2}}$$

$$< \frac{1}{n!}\left[\int_T \left[\int_{T^n} |k_n(t,s)|^2 ds \int_{T^n} |x(s)|^2 ds\right] dt\right]^{\frac{1}{2}}$$

$$= \frac{1}{n!}\left[\int_{T^n} |x(s)|^2 ds \int_T \int_{T^n} |k_n(t,s)|^2 dt ds\right]^{\frac{1}{2}}$$

$$= \frac{1}{n!}\left[\int_{T^n} |x(s_1)|^2 \ldots |x(s_n)|^2 ds\right]^{\frac{1}{2}} \left[\int_{T^{n+1}} |k_n(t,s)|^2 dt ds\right]^{\frac{1}{2}}$$

$$= \frac{1}{n!} \|x\|^n_{L^2(T)} \|k_n\|_{L^2(T^{n+1})}$$

Pour la convergence de $\sum^{\infty} \|u_n\|$, au lieu de supposer $\|x\|^n \|k_n\|$ borné $\forall n$, on imposera la majoration suivante beaucoup plus faible :

$\exists$ une constante K telle que pour chaque n, $\|k_n\| < K^n$

Dans ce cas, on a pour tout x :

$$\sum_{n=0}^{\infty} \|u_n\| \leq \sum_{n=0}^{\infty} \frac{1}{n!} \|x\|^n K^n = e^{K\|x\|}$$

Il y a donc convergence pour tout x; comme x est définie et continue sur le compact T, la convergence est alors uniforme en x.

Expression des noyaux en fonction de A :

Rappelons que (e_j) étant une base orthonormée de $L^2(T)$, les noyaux k_n étaient construits de la manière suivante dans le cas d'une fonctionnelle bilinéaire B : [1]

$$k(s_1,s_2) = \sum_{n,m=0}^{\infty} B(e_m,e_n)e_m(s_1)e_n(s_2)$$

Dans le cas des fonctionnelles multilinéaires $A_t^{(n)}$ (dérivée d'ordre n de la fonctionnelle A_t attachée à l'opérateur A), les noyaux se construisent d'une manière analogue :

$$k_n(t,s_1,..,s_n) = \sum_{k_1,..,k_n=0}^{\infty} A_t^{(n)}(e_{k_1},..,e_{k_n})e_{k_1}(s_1)..e_{k_n}(s_n)$$

$B(e_m,e_n)$ et $A_t^{(n)}(e_{k_1},..,e_{k_n})$ sont les coéfficients de Fourier des fonctions k et k_n relatifs aux bases orthonormées $\{e_m(s_1)e_n(s_2)\}$ ou $\{e_{k_1}(s_1)..e_{k_n}(s_n)\}$.

Expression de la majoration en fonction de A

La majoration a été obtenue pour k_n; on va la traduire en fonction de A ou de fonctionnelles liées à A.
On a :

$$\|k_n\|^2_{L^2(T^{n+1})} = \int_{T^{n+1}} \Big| \sum_{k_1,..,k_n=0}^{\infty} A_t^{(n)}(e_{k_1},..,e_{k_n})e_{k_1}(s_1)..e_{k_n}(s_n)\Big|^2 ds_1..ds_n dt$$

$$= \int_T \Big[\int_{T^n} \Big| \sum A_t^{(n)}(...)e_{k_1}(s_1)..e_{k_n}(s_n)ds_1..ds_n\Big|^2\Big]dt$$

D'aprés l'équation de Parseval :

$$\int_{T^n} \Big| \sum_{k_1..k_n=0}^{\infty} A_t^{(n)}(..)e_{k_1}(s_1)..e_{k_n}(s_n)\Big|^2 ds_1..ds_n$$

$$= \sum_{k_1..k_n=0}^{\infty} \Big|A_t^{(n)}(e_{k_1},..,e_{k_n})\Big|^2$$

et l'inégalité imposé à k_n devient :

$\exists$ une constante K telle que pour chaque n on ait :

$$\left[\int_T \left(\sum_{k_1..k_n=0}^{\infty} |A_t^{(n)}(e_{k_1},..,e_{k_n})|^2\right)dt\right]^{\frac{1}{2}} < K^n$$

Cette condition ne fait intervenir que des grandeurs fixées ou des fonctions directement liées à A.

P A R T I E 2

TRANSFORMATIONS NON LINEAIRES D'UN PROCESSUS ALEATOIRE

chap.1 : Densité des opérateurs H-S dans $L^p(X,\mu)$
approximation des transformations non linéaires d'un processus aléatoire.

chap.2 : Transformations non linéaires d'un processus défini par un système projectif de mesures ou par une mesure cylindrique.

CHAPITRE 1

DENSITE DES POLYNOMES H-S DANS $L^P(X,\mu)$

TRANSFORMATIONS NON LINEAIRES D'UN PROCESSUS ALEATOIRE

Soit $L^P(X,\mu)$ l'ensemble des applications A: $X \subset L^2(T) \to L^2(T)$, avec X compact muni d'une mesure μ, A étant mesurable et tel que :

$$\int_X \|A(x)\|^P d\mu(x) < \infty$$

On va élargir la classe des applications représentables par des polynomes H-S, classe qui jusque là se limitait à $C[X,L^2(T)]$, en montrant que l'ensemble des polynomes H-S est partout dense dans $L^P(X,\mu)$.

Ceci permet d'étudier les transformations non linéaires d'un processus aléatoire en les ramenant à des transformations intégrales à noyaux, celles-ci étant plus simples à étudier.
Eneffet si on interprète un processus comme une mesure sur un espace de fonctions (les réalisations sont de carré de module sommables presque sûrement si le processus est continu en moyenne quadratique), une transformation non linéaire (p-sommable) de ce processus se ramène à une transformation intégrale (polynome H-S).

On verra pour cette dernière qu'il est possible d'obtenir les caractéristiques probabilistes du processus transformé.

1.1. Approximation des applications p$^{\underline{e}}$ sommables de $L^2(T)$

Pour préciser le sens de 'A sommable', rappelons brièvement la construction de l'intégrale sur un espace mesuré $(\Omega,\mathcal{A},\mu)$:

soit A une fonctionnelle réèlle : $\Omega \to R$, avec $A>0$ et $\neq+\infty$
et soit $\mathcal{D}$: $0 \leqslant a_1 \leqslant \ldots \leqslant a_n \leqslant \ldots$ $(a_n \to \infty)$ une partition de R.

Les fonctions pour lesquelles les ensembles :

$$\left[a_j \leqslant A < a_{j+1} \right] = \{ \omega \in \Omega \ / \ a_j \leqslant A(\omega) < a_{j+1} \}$$

forment une σ-algèbre contenue dans $\mathcal{A}$ sont dites mesurables.

Si A est mesurable, son intégrale est par définition :

$$\int A d\mu = \lim_{\substack{\mathcal{D} \\ \sup(a_{j+1}-a_j)\to 0}} \left[\sum_{j=1}^{\infty} a_j \mu\left[a_j \leqslant A < a_{j+1}\right] \right] \in \bar{R}^+$$

Si l'espace d'arrivée de A n'est plus R mais un Banach, cette définition se généralise.[5]

Théorème de représentation

L'ensemble $\mathcal{H}$ des opérateurs H-S de X dans $L^2(T)$ est dense dans $L^p(X,\mu)$. Plus précisément :

Soit $A : X \subset L^2(T) \to L^2(T)$, avec X compact, Amesurable et tel que $\int_X \|A(x)\|^p d\mu(x) < +\infty$. Alors, $\forall \varepsilon > 0 \ \exists H \in \mathcal{H}$ tel que :

$$\|A - H\|_{L^p(X,\mu)} = \left[\int_X \|A(x) - H(x)\|^p d\mu(x) \right]^{\frac{1}{p}} \leqslant \varepsilon$$

Ce théorème se démontre aisément en remarquant que l'ensemble des fonctions continues est dense dans $L^p(X,\mu)$ et que $\mathcal{H}$ est dense dans $C[X,L^2(T)]$ (partie 1).

Démontrons le pour p=1 :

l'ensemble des fonctions continues est dense dans $L(X,\mu)$: $\forall \varepsilon>0$, $\exists f$ continue telle que :

$$\int_X \| A(x)-f(x) \| d\mu(x) < \frac{\varepsilon}{2}$$

l'ensemble $\mathcal{H}$ des H-S est dense dans $C[X,L^2(T)]$: $\forall \varepsilon>0 \; \exists H \in \mathcal{H}$ tel que :

$$\sup_{x \in X} \| A(x) - H(x) \| < \frac{\varepsilon}{2\mu(X)}$$

A mesurable et appartenant à $L(X,\mu)$, on a :

$$\int \|A(x)-H(x)\| d\mu(x) \leqslant \int \|A(x)-f(x)\| d\mu(x) + \int \|f(x)-H(x)\| d\mu(x)$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

- Pour p quelconque, la démonstration est analogue (densité des fonctions continues dans $L^p(X,\mu)$ et celle de $\mathcal{H}$ dans $C[X,L^2(T)]$, puis application de l'inégalité de Minkowski).

1.2. Approximation des transformations d'un processus aléatoire

Soit $(T,\mathcal{E})$ un intervalle de R muni de la σ-algèbre $\mathcal{E}=T\cap\mathcal{R}$ de ses parties boréliennes, et soit $(\Omega,\mathcal{A},P)$ un espace de probabilité.
Rappelons qu'on appelle fonction aléatoire réèlle mesurable une application mesurable χ de l'espace produit $(T\times\Omega,\mathcal{E}\otimes\mathcal{A})$ dans $(R,\mathcal{R})$; cette application χ transforme P en une mesure μ sur un espace de fonctions.

Si on interprète un processus aléatoire comme une mesure μ définie sur un espace de fonctions,(les réalisations du processus sont des fonctions appartenant presque sûrement à $L^2(T)$ si le processus est continu en moyenne quadratique [6]), l'approximation d'une transformation non linéaire du processus est ramenée à celle étudiée au paragraphe précédent, des opérateurs de $L^p(X,\mu)$.

Le cas genéral où le processus est défini par un système projectif de mesures sera discuté plus loin.
Ici, nous allons montrer que pour les transformations intégrales, on peut obtenir les caractéristiques probabilistes du processus transformé.

Soient Φ un ensemble de fonctions, Φ^* son dual, μ une mesure sur une σ-algèbre de Φ, $l(x)$ une fonctionnelle linéaire μ-mesurable pour tout $l\in\Phi^*$; on appelle fonctionnelle caractéristique (f.c.) la fonctionnelle :

$$\phi(l) = \int e^{il(x)}\mu(dx)$$

Soit un processus défini par μ, ou par les restrictions $\{\mu_{t_1..t_n}\}$ la fonctionnelle caractéristique est un moyen équivalent (utilisé pour les problèmes pratiques) pour décrire le processus (elle permet de déterminer les fonctions caractéristiques, et par là même, les répartitions finies).

Si le processus est continu en tant que fonction de t (par rapport à μ), la f.c. peut être définie par [7] :

$$\phi(g) = \int \exp\left[i\int x(t)dg(t)\right]\mu(dx)$$

Les processus considérés sont continus; on utilisera la deuxième définition pour les f.c.

Transformation intégrale d'un processus :

Soit la transformation du processus x(t) définie par :

$$y(t) = \int_{-\infty}^{\infty} \dots \int K[t,s_1,..,s_n,x(s_1),..,x(s_n)]ds_1..ds_n$$

Soient :

$$\phi_x[g(.)] = E\left[\exp\left\{i\int_{-\infty}^{\infty} x(t)dg(t)\right\}\right]$$

la f.c. de x(t), et $\phi_y[g(.)]$ celle de y(t); (E=espérance math.);

Posons :

$$u_\lambda(g,g_1) = E\exp\left[i\int_{\infty}^{\infty} x(t)dg(t) + i\lambda\int_{\infty}^{\infty} y(t)dg_1(t)\right]$$

On supposera que la fonction K admet la représentation :

$$K(t,s_1,..,s_n,x_1,..,x_n) = \int..\int\left[e^{i\Sigma r_k x_k} - 1\right]N_{t,s_1,..,s_n}(dr_1..dr_n)$$

où pour $t,s_1,..,s_n$ fixés, $N_{t,s_1,..,s_n}$ est une mesure sur R^n.

On montre [8] que la fonction $u_\lambda(g,g_1)$ satisfait l'équation :

$$\frac{d}{d\lambda}\, u_\lambda(g,g_1) = i\int..\int u_\lambda(g,g_1+\Sigma r_k\varepsilon_{s_k}) - u_\lambda(g,g_1)\, N_{t,s_1,..,s_n}(dr_1..dr_n) \times ds_1..ds_n dg_1(t)$$

avec : $\begin{cases}\varepsilon_s(t)=1 \text{ pour } t> s\\ \varepsilon_s(t)=0 \qquad t\leqslant s\end{cases}$

u_λ obtenu à partir de cette équation fournit la f.c. du processus transformé:

$$\begin{cases} u_\lambda(g,0) = u_0(g,g_1) = \phi_x(g)\\ u_\lambda(0,g_1) = \phi_y(\lambda g_1)\end{cases}$$

-L'étude numérique proprement dite de l'approximation de A sera faite en parti

CHAPITRE 2

TRANSFORMATIONS NON LINEAIRES D'UN PROCESSUS DEFINI PAR UN SYSTEME PROJECTIF DE MESURES

On va discuter les divers cas où les transformations non linéaires d'un processus se ramènent à des transformations intégrales du type H-S.

On a vu que si A est non linéaire et défini sur un espace mesuré on peut toujours l'approximer par un opérateur H-S, pourvu que A soit mesurable et $\tilde{p}$ sommable. Ce résultat s'applique évidemment au cas où A est défini sur un processus et si celui-ci est interprété comme une mesure sur un espace fonctionnel.

En général, un processus est défini par un système projectif de mesures $\{ \mu_{t_1..t_n} \}$. La limite projective, lorsqu'elle existe, est la mesure μ à laquelle on assimile le processus.

On discutera d'abord les divers cas d'existence de cette limite, cas dont l'étude se ramène à celle du chapitre précédent. Pour cela on rappelera brièvement la définition d'un système projectif de mesures et son utilisation pour la définition d'une fonction aléatoire.

Si la limite projective n'existe pas, autrement dit le processus n'est pas défini par une vraie mesure, problème peu intéressant du point de vue probabiliste mais théoriquement possible(correspondance bijective entre les mesures cylindriques sur un Hilbert séparable et les fonctions aléatoires linéaires réèlles sur ce même espace), le problème devient très difficile car les propriétés topologiques de A étaient liées à l'existence d'une vraie mesure μ.

Pour généraliser le résultat sur la densité des H-S dans $L(X,\mu)$ à $L(X, \{ \mu_{t_1..t_n} \})$, espace des opérateurs sommables vis à vis d'un système

projectif défini par une mesure cylindrique, il faut définir ce qu'on entend par A sommable et par approximation de A; cela revient à définir l'intégrale de A vis à vis d'un système projectif de mesures et une topologie dans $L(X, \{ \mu_{t_1..t_n} \})$.

L'intégrale peut toujours être définie (par une méthode analogue à celle de Wiener; on définira des sommes par discrétisation de A le long d'une ligne brisée , et on prendra la limite de ces sommes selon le filtre des complémentaires des parties finies.).

Toutefois, la difficulté apparait lorsqu'on parle d'approximation de A car on ne connait pas de topologie sur l'espace $L(X, \{ \mu_{t_1..t_n} \})$; le seul cas mais simple puisqu'il nous ramène à R^n, où on peut parler de limite, est celui des opérateurs A dépendant d'un nombre fini de paramètres $x(t_1),...,x(t_n)$.

2.1. Système projectif de mesures

Un espace topologique séparé est dit complètement régulier s'il est homéomorphe à un sous espace d'un espace compact.

définition :

Soient (X_i) une famille d'espaces complètement réguliers où i parcourt un ensemble ordonné d'indices I.
Pour tout couple (i,j) tels que $i \leqslant j$ on suppose donnée une application continue :

$$f_{ij} : X_j \longrightarrow X_i$$

telle que :

- $f_{ik} = f_{ij} \circ f_{jk}$ si $i \leqslant j \leqslant k$
- f_{ii} = identité dans X_i
- Pour tout i, μ_i est une mesure sur X_i de masse totale 1, satisfaisant à :

$$\mu_i = f_{ij}(\mu_j) \qquad \text{pour tout couple } (i,j) \quad i \leqslant j$$

Le système (X_i, f_{ij}, μ_i) satisfaisant ces trois conditions est un système projectif de mesures.

Problème :

Appellons Z la limite projective du système (X_i, f_{ij}), c'est à dire le sous espace fermé du produit des X_i tel que :

$$x = (x_i) \in Z \quad \leftrightarrow \quad x_i = pr_i(x) = f_{ij}(x_j) = f_{ij}(pr_j(x)) \quad i \leqslant j$$

On cherche une mesure μ sur Z telle que pour tout i : $\mu_i = f_i(\mu)$ où f_i est l'application canonique de Z dans X_i définie par la restriction de pr_i à Z.
Cette mesure, lorsqu'elle existe, sera notée $\mu = \varprojlim \mu_i$, et dite limite projective.

Exemple : mesure cylindrique [9] [10]

Soit E un espace localement convexe séparé, et soit $\mathcal{N}$ la famille des sous espaces fermés de E de codimension finie; c'est une base de filtre.

Soit $N \in \mathcal{N}$; on designe par E_N l'espace E/N : c'est un espace

vectoriel de dimension finie.

Si $N_1 \subset N_2$, on a une application linéaire canonique de E_{N1} sur E_{N2} ; on l'appellera $\Pi_{N_2N_1}$; les applications $\Pi_{N_2N_1}$, $N_1 \subset N_2$ satisfont la relation de cohérence :

$$N_1 \subset N_2 \subset N_3 \qquad \Pi_{N_3N_1} = \Pi_{N_3N_2} \circ \Pi_{N_2N_1} .$$

Les E_N , $N \in \mathcal{N}$ forment un système projectif d'espaces vectoriels (relativement aux $\Pi_{N_1N_2}$).

On appelle mesure cylindrique ou promesure sur E un système projectif $(E_N,\mu_N)_{N \in \mathcal{N}}$ où μ_N est une probabilité de Radon sur E_N.

La proposition suivante permet de remplacer les quotients de dimension finie par n'importe quel espace de dimension finie :

Proposition :

Se donner une mesure cylindrique sur E revient à se donner sur tout espace vectoriel de dimension finie X et toute application linéaire continue f de E dans X une probabilité de Radon sur X, μ_f, les μ_f satisfaisant la relation de cohérence :

$$E \xrightarrow{f} X \xrightarrow{g} Y \qquad \mu_{g\circ f} = g(\mu_f)$$

Fonction aléatoire et système projectif de mesures [11]

La donnée d'une fonction aléatoire (f.a.) est équivalente à celle d'une famille de variables aléatoires (v.a.) $X_t(\omega)$ dépendant d'un paramètre $t \in T$, ce qui est encore équivalent à la donnée d'une famille d'espaces mesurables $\{(\Omega_t,\mathcal{A}_t);t \in T\}$.

Pour décrire les relations mutuelles entre une infinité de variables aléatoires, on se donne en général toutes les relations probabilistes de chaque ensemble fini de valeurs de la f.a. : $X_t(\omega),\ldots,X_{t_n}(\omega)$, ce qui revient à se donner pour tout n-uple $(t_1,..,t_n)$ la loi de probabilité $P_{t_1,..,t_n}$ de la variable aléatoire :

$$(X_{t_1}(\omega),..,X_{t_n}(\omega)) \quad \text{définie sur } \Big(\prod_1^n \Omega_{t_i} , \bigotimes_1^n \mathcal{A}_{t_i} \Big).$$

On cherche ensuite à définir une probabilité P sur l'espace mesurable $\{ \prod_T \Omega_t , \otimes \mathcal{A}_t \}$ des trajectoires, cette probabilité devant être telle que ses restrictions aux σ-algèbres $\bigotimes_1^n \mathcal{A}_{t_i}$ soient égales aux probabilités $P_{t_1..t_n}$ données; ceci est traduit par la condition de compatibilité qui fait que la famille $\{ P_{t_1..t_n} \}$ forme un système projectif de mesures :

Pour qu'il existe une fonction d'ensembles P définie sur $[\prod_T \Omega_t , \bigcup (\bigotimes_1^n \mathcal{A}_{t_i})]$ dont les restrictions aux σ-algèbres $\bigotimes_1^n \mathcal{A}_{t_i}$ soient les $P_{t_1..t_n}$, il faut et il suffit que si $T_1 \subset T_2$ (T_1 et T_2 étant deux sous ensembles finis de T) la restriction de P_{T_2} à $\bigotimes_{T_1} \mathcal{A}_t$ soit P_{T_1}.

Existence de la limite projective :

- cas d'une fonction aléatoire

En général on impose la condition suivante (propriété d'approximation) pour avoir l'existence de la limite projective :

pour tout $t \in T$, il existe une sous classe compacte $\mathcal{C}_t$ de $\mathcal{A}_t$ telle que :

$$P_{(t)}(A) = \sup [P_{(t)}(C) , C \in \mathcal{C}_t , C \subset A] \quad \text{pour tout } A \in \mathcal{A}_t$$

Avec cette condition, il existe une probabilité P unique sur l'espace produit $(\prod_T \Omega_t , \bigotimes_T \mathcal{A}_t)$ qui prolonge chacune des probabilités $P_{t_1;..t}$

- cas d'un système projectif

Soit (X_i , f_{ij} , μ_i) un système projectif de mesures; si les espaces X_i sont compacts, $\varprojlim \mu_i$ existe et est unique.

Sinon, dans le cas général, il y a le théorème de Prokorov :

Soient (X_i , f_{ij} , μ_i) un système projectif de mesures sur des espaces complètement réguliers, E un espace complètement régulier et pour tout i une application :

$$f_i : E \to X_i \quad \text{telle que} \quad f_i = f_{ij} \circ f_j$$

Pour qu'il existe une mesure μ sur E telle que $f_i(\mu)=\mu_i \ \forall i$, il faut et il suffit que pour tout $\varepsilon > 0$ donné, il existe un compact $K \subset E$ tel que $\forall i$:

$$\mu_i(f_i(K)) \geqslant 1 - \varepsilon$$

2.2. Transformations non linéaires d'un processus défini par un système projectif :

2.2.1. Cas se ramenant à une vraie mesure

Dans tous les cas où on impose les conditions précédentes au système projectif (propriétés d'approximation, conditions du théorème de Prokorov), on arrive à une vraie mesure pour définir la f.a. et on est ainsi ramené au cas déjà étudié d'approximation d'opérateurs définis sur un espace mesuré.

Un cas où on s'y ramène aussi par utilisation du résultat sur les mesures sur un produit de compacts, est celui où A est défini sur une partie X de $L^2(T)$ avec $x \in X$, $t \in T$, et $x(t) \in [a,b]$. L'ensemble de toutes ces fonctions de T dans $[a,b]$ est $[a,b]^T$, donc un compact puisque produit d'une famille de compacts (Tychonoff); il suffit à ce moment pour avoir une mesure sur X de se donner des mesures $\mu_{t_1..t_n}$ sur $[a,b]^n$ formant un système projectif.

Souvent, il n'est même pas nécessaire d'imposer la condition de système projectif car c'est inclus dans la construction même.

Ainsi par exemple, soient X_i des compacts et μ_i une mesure sur X_i; on construit une mesure μ_J sur le produit :

$$X_J = \prod_{i \in J} X_i \qquad\qquad J \in \mathcal{F}(I) \quad \text{(parties finies)}$$

en prenant le produit tensoriel :

$$\mu_J = \bigotimes_{i \in J} \mu_i$$

μ_J est une mesure sur X_J; si on suppose $\mu_i(X)=1$, alors la famille (μ_J) est

un système projectif de mesures :
En effet, $\forall J,K$ avec $J \subset K$, la projection pr_{JK}: $X_K \to X_J$, est telle que $pr_{JK}(\mu_K)=\mu_J$.

Ainsi, à partir de mesures sur les X_i, on a une mesure ($\varprojlim$) sur le produit infini.

2.2.2. Cas où la limite projective n'existe pas; définition de $L[X, \mu_{t_1..t_n}]$

Le cas où le processus n'est pas défini par une vraie mesure est rarement étudié en calcul des probabilités, mais il peut arriver théoriquement qu'une f.a. découle d'une mesure cylindrique comme le montre le théorème de L. Schwartz :

Il y a correspondance bijective entre les mesures cylindriques sur H (H espace de Hilbert séparable) et les f.a. linéaires réèlles sur H à une isonomie près (il lui correspond une classe de f.a. isonomes).

(Rappelons que deux v.a. à valeurs dans H : χ: $(\Omega,P)\to H$ et χ' : $(\Omega',P')\to H$ sont isonomes si $\chi(P)= \chi'(P')$, autrement dit elles définissent sur H la même loi de probabilité.
Une f.a. χ_t définie sur T et à valeurs dans H (application χ_t: $t \in T \to$ v.a. à valeurs dans H, T étant un espace vectoriel) est dite linéaire si $\chi_{t_1+t_2} = \chi_{t_1} + \chi_{t_2}$).

intégrale d'un opérateur par rapport à un système projectif de mesures

Si le système projectif définissant le processus n'a pas de limite projective, l'intégrale peut toujours être définie par analogie avec celle de Wiener. [12]

Soit A un opérateur défini sur X, X étant l'espace des fonctions de T dans R, et T un intervalle réèl.

On appellera ensemble cylindrique à base pavé ou quasi-intervalle les ensembles ayant la forme suivante :

$$B = \{ x(t) \in X : \alpha_1 < x(t_1) \leqslant \beta_1, \ldots, \alpha_n < x(t_n) \leqslant \beta_n \}$$

Ces ensembles forment un semi-anneau.

Si pour chaque sous ensemble de R^n, $(t_1,..,t_n)$ étant fixés, on a une mesure $\mu_{t_1 \ldots t_n}$, on peut obtenir une mesure du quasi-intervalle B en posant :

mesure de B dans X = mesure de sa base dans R^n

ou :

$$\mu(B) = \mu_{t_1..t_n}\left[\alpha_j < \xi_j \leqslant \beta_j \;\; j=1,..,n \right]$$

Si on suppose les $\mu_{t_1..t_n}$ compatibles, au sens :

$$\mu_{t_1..t_n}\left[\alpha_j < x(t_j) \leqslant \beta_j \;\; j=1,..,n \right] = \mu_{t_1..t_{n+1}}\left[\alpha_j < x(t_j) \leqslant \beta_j \;\; j=1,..,n \quad -\infty < x(t_{n+1}) \leqslant \infty \right]$$

on a alors un système projectif de mesures. [13]

On va définir l'intégrale de A par rapport à ce système projectif; on pose :

$$x_q(t) = x(t_k) \quad \text{pour} \quad t_k \leqslant t < t_{k+1} \quad k=1,..,n \quad q=\text{subdivision de T}$$

Pour chaque q fixé, la valeur :

$$A_q[x(t)] = A[x_q(t)]$$

de A définit une nouvelle fonctionnelle dépendant des valeurs de x(t) seulement aux points $t_1,\ldots,t_n$:

$$A_q[x(t)] = A_q(x_1,\ldots,x_n) \qquad (x_k = x(t_k))$$

Supposons que A_q est intégrable par rapport à $\mu_{t_1..t_n}$ dans R^n et posons:

$$I_q(A) = \int_{R^n} A_q(x_1,..,x_n)d\mu_{t_1..t_n}$$

On appellera intégrale de A par rapport au système projectif la limite (selon le filtre des complémentaires des parties finies de T) suivante lorsqu'elle existe :

$$I(A) = \lim_q I_q(A)$$

$\mu = \mu(q,\alpha_j,\beta_j) = \{\mu_{t_1..t_n}\}$

(μ est la mesure cylindrique qui définit le système projectif $\{\mu_{t_1..t_n}\}$)

On notera $L(X,\{\mu_{t_1..t_n}\})$ l'ensemble des opérateurs A tels que $|I(A)| < +\infty$.

Pour parler d'approximation de A dans cet espace, il faudrait y définir une topologie.

L'étude théorique s'arrêtera là car on ne sait pas en général définir une topologie dans $L[X,\{\mu_{t_1..t_n}\}]$, sauf dans certains cas simples. Cela a lieu par exemple, si on définit la convergence d'une suite de fonctionnelles $A_m[x(t)]$ vers zéro comme suit :

toutes les fonctionnelles A_m dépendent des valeurs de x(t) en un même ensemble de points, si à l'infini elles tendent dans un certain sens uniformément vers zéro, et si pour chaque x(t) elles tendent vers 0 lorsque $m \to \infty$; mais l'espace $L[X,\{\mu_{t_1..t_n}\}]$ est dans ce cas trés limité puisqu'on revient à des fonctions numériques intégrables dans R^n.

CONCLUSION :

Ainsi on peut représenter des applications non linéaires continues, analytiques ou $p^{ème}$ sommables par des opérateurs polynomes H-S.

Cette étude a été menée par analogie avec la représentation des fonctions numériques continues, analytiques ou $p^{ème}$ sommables par les polynomes algébriques usuels.

P A R T I E 3

APPLICATION A L'IDENTIFICATION DES SYSTEMES

chap.1 : Représentation des systèmes non linéaires
utilisation d'opérateurs H-S orthogonaux; cas multidimensionnel; systèmes aléatoires.

chap.2 : Entrée déterminante et classes déterminables

CHAPITRE 1

REPRESENTATION DES SYSTEMES NON LINEAIRES

Un système est un mécanisme qui à une fonction x(s) associe une autre fonction y(t) (ou à un processus $x(s,\omega)$ un autre processus $y(t,\omega)$).

On peut traduire cette correspondance par un opérateur A.

On a montré que A pouvait, sous certaines conditions (A continu, analytique ou intégrable, cette dernière concernant surtout les processus aléatoires), être représenté par un polynome H-S.

Les hypothèses sur A sont évidemment celles faites sur le système; on peut les considerer comme physiquement faibles et satisfaites par une grande classe de systèmes non linéaires.

Le modèle représentant le système est un polynome H-S; la sortie s'écrit :

$$y(t) = A\left[x(s)/s \leqslant t\right] = \sum_n \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

1.1. Utilisation d'opérateurs H-S orthogonaux

Au lieu de représenter le système par un H-S ordinaire, on peut utiliser des opérateurs H-S orthogonaux (cas de l'espace $L^2[X,\mu]$ seulement puisque pour $L^p[X,\mu]$ $p\neq 2$, il n'y a pas de produit scalaire mais un produit de dualité).

Une telle représentation est intéressante pour les systèmes aléatoires. En effet si la sortie du système est :

$$y(t) = \sum_n \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n = \sum_n H_n(t)$$

et l'orthogonalité définie par (E= espérance mathématique)

$$E[H_i(t)H_j(t)] = \begin{cases} 0 \text{ si } i\neq j \\ E[H_i^2(t)] \text{ si } i=j \end{cases}$$

L'avantage est que, si un terme $H_k(t)$ de la représentation orthogonale obtenue décrit une propriété statistique liant l'entrée et la sortie, propriété liée à l'espace de Hilbert $L^2[X,\mu]$, ce terme la décrit complètement puisqu'il est orthogonal (donc indépendant) à tous les autres termes de la représentation.

Soit X compact dans $L^2(T)$ et muni d'une mesure μ :
Pour remplacer le système d'opérateurs H-S complet dans $L^2[X,\mu]$ par un système complet de polynomes H-S orthogonaux, on va appliquer la définition de l'orthogonalité dans $L^2[X,\mu]$ et le procédé d'orthogonalisation de Schmidt.

Le produit scalaire est défini dans $L^2[X,\mu]$ par :

$$(F,G) = \int_X F(x)G(x)d\mu(x)$$

F et G $\in L^2[X,\mu]$ sont orthonormaux si :

$$(F,G)=0 \quad \text{et} \quad \int_X [F(x)]^2\mu(dx) = \int_X [G(x)]^2\mu(dx) = 1$$

Si $H_0, H_1, .., H_n, ..$ est le système initial, le système orthonormé $H'_0, H'_1, ..., H'_n, ..$ s'obtient par détermination des coéfficients a_{ij} dans la transformation linéaire (détermination à partir des conditions d'orthogonalité) :

$$H'_0 = a_{00} H_0$$
$$H'_1 = a_{11} H_1 + a_{10} H_0$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$H'_n = a_{nn} H_n + \ldots + a_{n0} H_0$$

Le théorème de représentation obtenu avec (H_i) demeure en utilisant (H'_i) .

1.2. Système à plusieurs entrées

L'entrée $\bar{x} = [x^{(1)}(t), .., x^{(n)}(t)]$ est définie dans $\bar{X} = \prod_{i=1}^{n} X_i$, avec $x^{(i)}(t) \in X$.

Si $(H^i_{k_i}[x^{(i)}]\ k_i = 1,2,..)$ est un système d'opérateurs H-S dense dans $L^p[X_i, \mu_i]$ on obtient un système dense dans $L^p(\prod_{i=1}^{n} X_i, \mu = \otimes \mu_i)$ en prenant les produits tensoriels :

$$H_{0\ldots0}[\bar{x}] = H^1_0[x^{(1)}] \times \ldots\ldots \times H^n_0[x^{(n)}]$$
$$H_{1,..0}[\bar{x}] = H^1_1[x^{(1)}] \times \ldots\ldots \times H^n_0[x^{(n)}]$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$H_{k_1..k_n}[\bar{x}] = H^1_{k_1}[x^{(1)}] \times \ldots\ldots \times H^n_{k_n}[x^{(n)}]$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

On obtient ainsi l'ensemble des opérateurs :

$$\left[H_{k_1,..,k_n}[\bar{x}] = H^1_{k_1}[x^{(1)}] \times \ldots \times H^n_{k_n}[x^{(n)}] \right]_{\substack{k_1=0,1,\ldots \\ k_2=0,1,\ldots \\ \ldots\ldots}} \qquad (1)$$

L'ensemble (1) est dense dans $L^2\left[\prod X_i\ ,\ \bigotimes \mu_i\right]$ et orthonormé si le système :

$$\left[H^i{}_{k_i}(x^{(i)}) \right]_{k_i=0,1,2,..}$$

l'est dans $L^2\left[X_i,\mu_i\right]$.

$$\int_{\bar{X}} H_{k_1..k_n}(\bar{x})\ H_{j_1..j_n}(\bar{x})\ d\mu(\bar{x}) =$$

$$= \int_{\bar{X}=\prod X_i} H^1_{k_1}(x^{(1)})..H^n_{k_n}(x^{(n)})H^1_{j_1}(x^{(1)})..H^n_{j_n}(x^{(n)})d\mu(x^{(1)})..d\mu(x^{(n)})$$

$$= \int_{X_1} H^1_{k_1}(x^{(1)})H^1_{j_1}(x^{(1)})d\mu(x^{(1)}) \times \int_{X_2}...\times \int_{X_n} H^n_{k_n}(x^{(n)})H^n_{j_n}(x^{(n)})d\mu(x^{(n)})$$

$$= \delta_{k_1j_1}\times....\times\delta_{k_nj_n} \qquad (\delta = \text{symbole de Kronecker})$$

D'où le théorème de densité pour le cas multidimensionnel :

Théorème :

Le système $\left\{ H_{k_1..k_n}(\bar{x}) = \prod_{i=1}^{n} H^i_{k_i}\left[x^{(i)}(t)\right] \right\}_{\substack{k_i=0,1,.. \\ i=1,..,n}}$

est dense dans $L^2\left[\bar{X},\mu\right]$.

Autrement dit, si $A \in L^2\left[\bar{X},\mu\right]$, alors :

$$\lim_{N\to\infty} \int_{X_1}...\int_{X_n} \left\| A(\bar{x}) - \sum_{k_1,..,k_n=0}^{N} c_{k_1..k_n} H_{k_1..k_n}(\bar{x}) \right\|^2 d\mu(x^{(1)})..d\mu(x^{(n)}) = 0$$

où les $c_{k_1..k_n}$ sont les coéfficients de Fourier :

$$c_{k_1..k_n} = \int_{X_1}...\int_{X_n} A(\bar{x}) H_{k_1..k_n}(\bar{x}) d\mu(x^{(1)})..d\mu(x^{(n)})$$

Ce dernier résultat peut s'appliquer par exemple à une entrée bruitée en écrivant :

$$A[\bar{x}(t)] = A[x(t),\xi(t)] \qquad \text{si } \bar{x} = x(t) + \xi(t)$$

Dans ce cas :

$$A[\bar{x}(t)] = \sum_{i,j} c_{ij} H_i^{(1)}[x] H_j^{(2)}[\xi]$$

où $H_i^{(1)}$ et $H_j^{(2)}$ sont des H-S.

1.3. Systèmes aléatoires

Jusque là, les aléas etaient dans l'entrée et la sortie et non à l'intérieur du système; L'opérateur A était déterministe mais agissait sur des fonctions aléatoires.

Il y a des cas où l'opérateur est lui-même aléatoire (systèmes ayant à l'intérieur un bruit de fond, systèmes self-oscillants,..).

Pour de tels systèmes, à une réalisation x(t) de la f.a. entrée X(t), il correspond une f.a. Y(t) :

$$Y(t) = Ax(s)$$

On supposera évidemment qu'il existe une relation entre la f.a. entrée X(t) et la f.a. sortie Y(t); autrment dit, si on appelle E_y l'évènement consistant en la correspondance entre une réalisation donnée x de l'entrée et un ensemble de réalisations des sorties, la probabilité conditionnelle $P_x(E_y)$ est non nulle.

Ainsi, à une réalisation spécifiée x(t) de l'entrée, il correspond une répartition de probabilité sur l'ensemble des réalisations des sorties, repartition caractérisée par la probabilité $P_x(E_y)$.

L'opérateur aléatoire A peut être décrit par une famille d'opérateurs déterministes $(A_\omega)_{\omega \in \Omega}$, Ω etant l'espace des épreuves pour A.

D'aprés l'étude précédente, on a pour chaque ω fixé :

$$y(t) = A[x(s)] = A_\omega[x(s)] = \sum_n \int_{T^n} k_n(\omega,t,s_1,..,s_n)x(s_1)...x(s_n)ds_1..ds$$

Les aléas de l'opérateur se manifestent par des aléas sur les fonctions noyaux.
Donc, ponctuellement (pour ω fixé), l'opérateur A est représentable par un opérateur H-S à noyaux aléatoires.
(l'espace aléatoire de A et celui de l'entrée sont indépendants.)

Pour identifier de tels systèmes, on détermine en géneral une estimation $\hat{A}$ de A, ce qui revient à estimer les noyaux k_n. L'estimation $\hat{A}$ sera recherchée sous forme d'opérateurs H-S, mais déterministes; $\hat{A}$ étant obtenu, on est ramené au cas déjà étdié d'opérateurs déterministes agissant sur des fonctions aléatoires.

Ainsi, à la f.a. entrée X(s), il correspondra une f.a. sortie estimée $Y^*(t)$ définie par :

$$Y^*(t) = \hat{A}X(s)$$

Les noyaux k_n de $\hat{A}$ seront déterminés, on le verra plus loin, comme pour les cas précédents en minimisant l'écart quadratique moyen :

$$\min_{k_i} E[Y(t) - Y^*(t)]^2$$

On a vu qu'il était intéressant d'utiliser pour $\hat{A}$ un développement orthogonal :

$$Y^*(t) = \sum_k Y_k(t)$$

avec $Y_n(t) = \int k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$

$$\text{et} \quad E[Y_i(t)Y_j(t)] = \begin{cases} 0 & i \neq j \\ E(Y_i^2(t)) & \text{si } i=j \end{cases}$$

CHAPITRE 2

ENTREE DETERMINANTE POUR L'IDENTIFICATION

CLASSES DETERMINABLES

Le problème de l'identification d'un système est celui de la determination de son opérateur A (ou de l'approximation $\hat{A}$ de A).
$\hat{A}$ fournit à l'instant t la sortie (on dira sortie calculée ou sortie du modèle) représentée par :

$$y_c(t)=(\hat{A}x)(t)= \hat{A}\left[x(s)/s \leqslant t\right]= \sum_n \int_{T^n} k_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1.ds_n$$

En pratique, on prend les N premiers termes de ce développement et on détermine les noyaux k_n de façon à minimiser un critère mesurant l'écart entre la sortie vraie ou mesurée $A\left[x(s)\right]$ et la sortie calculée $\hat{A}\left[x(s)\right]$.

On utilisera le critère des moindres carrés avec comme données des valeurs mesurées de certaines entrées et des sorties correspondantes.

La détermination de $\hat{A}$ devient alors du calcul formel avec toutefois un problème : le choix des entrées mesurées utilisées dans le critère d'identification.

En effet, si $\hat{A}$ a été déterminé par une fonction entrée x(s) par exemple, l'écart entre la sortie vraie $A\left[x(s)\right]$ et la sortie du modèle $\hat{A}\left[x(s)\right]$ est petit. Le modèle $\hat{A}$ étant maintenant déterminé et fixé à partir de x(s), est essayé sur une autre entrée x'(s). Si on compare les résultats de $A\left[x'(s)\right]$ et $\hat{A}\left[x'(s)\right]$ (valeurs de la sortie vraie et de la sortie calculée pour la seconde entrée x'(s) ou 'entrée test'), on les trouve généralement totalement différents.

Autrement dit, $\hat{A}$ risque d'approximer A seulement pour les entrées ayant servies à sa détermination.

Il est évident que l'approximation et la précision seront d'autant meilleures que le nombre de mesures d'entrées-sorties utilisées dans le critère est grand, et les entrées utilisées variées.
En pratique, on est souvent limité et il faudra faire un choix de ces entrées déterminantes.

Ce chapitre étudie le choix de ces entrées determinantes pour l'identification; il sera plutôt théorique et fera appel à des méthodes d'analyse spectrale. son but n'est pas de résoudre le problème, (la solution empirique consiste à prendre un grand nombre de fonctions entrées, ou un processus aléatoire), mais de montrer les possibilités d'identification en utilisant un nombre minimum de fonctions entrées.

2.1. Etude de l'entrée déterminante

solution pratique

Ce problème fut considéré par Wiener [14] qui utilisa comme entrée déterminante une fonction aléatoire, plus spéciallement un bruit blanc† Gaussien (dérivée par rapport au temps d'un processus de Wiener). cela parait naturel car lors de son évolution, le processus entrée prend des valeurs appartenant à plusieurs fonctions déterministes, celles ci étant des réalisations du processus.

L'utilisation d'un processus aléatoire comme entrée fournit ainsi le même résultats que plusieurs fonctions entrées déterministes.

Ceci correspond bien à la solution empirique consistant à prendre beaucoup de mesures d'entrées-sorties, et des entrées variées. Il est donc intéréssant d'utiliser un processus trés irrégulier.

Problème

Est-il possible d'identifier des systèmes à partir d'une seule fonction entrée déterministe?
Une certaine approche de ce problème va être faite en utilisant la notion de classes déterminables introduite en [16]

† Généralement les physiciens appellent 'bruit blanc' une fonction aléatoire réèlle X_t d'espérance mathématique nulle et dont la fonction de corrélation $\Gamma(t_1,t_2) = E[X_{t_1} X_{t_2}]$ est donnée formellement par :

$$\Gamma(t_1,t_2) = \delta(t_1 - t_2)$$

où δ est la distribution de Dirac.

2.2. Identification à partir d'une seule fonction entrée

L'approche du problème se fera à partir de la notion de 'classes déterminables'.
On va caractériser les situations où l'identification est possible.
On ne considerera que les systèmes représentables par des opérateurs H-S.

Les fonctions entrées $x \in X \subset L^2(T_1)$, les sorties $y \in L^2(T)$, T_1 et T_2 étant des intervalles réèls.

Soit $H : X \to L(T_2)$ un opérateur H-S; on notera $\mathcal{H}$ l'espace vectoriel normé de tels opérateurs H ($\|H\|$ et $|||H|||$ désigneront respectivement la norme du sup et la norme H-S).

définition 1 :

Soit $\varepsilon > 0$ et $\mathcal{D} \subset \mathcal{H}$; on dira que $\mathcal{D}$ est une classe ε-déterminable si $x \in X$, un ensemble fini de fonctionnelles continues à valeurs réèlles $m_i(y)$ $i=1,..,N$, bornées sur les ensembles bornés, et une fonction continue $\hat{H}$ de R^N dans $\mathcal{H}$ tels que :

$$\| H - \hat{H}[m_1(Hx),...,m_N(Hx)] \| \leq \varepsilon \qquad \forall H \in \mathcal{D} \qquad (N=N(\varepsilon))$$

Signification pratique :

Tous les systèmes considérés sont décrits par des opérateurs H-S; H est l'opérateur du système, x(t) est l'entrée et Hx la sortie.
La classe ε-déterminable $\mathcal{D}$ est la classe des opérateurs appartenant à $\mathcal{H}$ pouvant être approximés à ε près, par utilisation d'une seule fonction entrée déterminante x(t).

$m_1(Hx),..,m_N(Hx)$ seront les mesures faites sur la sortie Hx à des instants fixés $t_1,..,t_N$.

En pratique, à partir des mesures $m_j(Hx)$, on détermine un opérateur (fixé une fois pour toutes par ces mesures) $\hat{H}[m_1(Hx),..,m_N(Hx)]$ qui approxime H; autrement dit, au vecteur $[m_1(Hx),..,m_N(Hx)] \in R^N$, on associe $\hat{H} \in \mathcal{H}$.

$\hat{H}$ approxime H en norme ($\|H-\hat{H}\| \leq \varepsilon$); donc, bien que déterminé par x, $\hat{H}$ approxime H pour des fonctions autres que x.

définition 2 :

On dit que $\mathcal{D}$ est une classe déterminable si $\mathcal{D}$ est une classe ε-déterminable pour tout $\varepsilon>0$.
($x, N, \{m_i\}$, et $\hat{H}$ peuvent tous dépendre de ε)

définition 3 :

On dit que $\mathcal{D}$ est déterminable par rapport à x si $\mathcal{D}$ est ε-déterminable pour tout $\varepsilon> 0$, mais toujours avec le même x.

Remarquons que dans la définition 2, $\mathcal{D}$ est déterminable pour tout $\varepsilon>0$, mais l'entrée x utilisée dépendait de ε et changeait ainsi avec ε. La définition 3 impose une certaine uniformité.

Quand un ensemble $\mathcal{D}$ d'opérateurs peut-il être déterminable?

Supposons que $\mathcal{D}$ soit une partie bornée de $\mathcal{H}$ ($\forall H \in \mathcal{D}, \|H\| \leq C$); on montre que : [15]

Proposition :

$\mathcal{D}$ est une partie déterminable si et seulement si :

1. $\mathcal{D}$ est précompacte ($\overline{\mathcal{D}}$ compacte)
2. Pour tout $\varepsilon>0$, il existe un $x=x(\varepsilon) \in X$ tel que pour tous les H_1 et $H_2 \in \mathcal{D}$ satisfaisant $\|H_1 - H_2\| > \varepsilon$, on ait :

$$\|H_1x - H_2x\| >0$$

Opérateurs identifiables à partir d'une seule entrée x

Soit $\mathcal{H}$ l'ensemble des opérateurs H définis par:

$$y(t) = (Hx)(t) = \sum_{n=0}^{N} \int_{T_1^n} h_n(t,u_1,..,u_n)x(u_1)..x(u_n)du_1..du_n$$

Par le changement de variables $s_i=t-u_i$, et en posant :

$$k_n(t,s_1,..,s_n) = h_n(t,t-s_1,..,t-s_n)$$

on obtient :

$$y(t) = \sum_{n=0} \int_{T^n} k_n(t,s_1,..,s_n)x(t-s_1)..x(t-s_n)ds_1..ds_n \qquad (1)$$

$$t \in T_2 \quad s_i \in T$$

hypothèses :

Dans (1) on peut supposer le terme de degré zéro nul car il l'est souvent; sinon, il peut être déterminé d'avance à partir des conditions initiales et soustrait de la série.

On supposera que le système ne varie pas avec le temps (stationnaire) :

$$k_n(t,s_1,..,s_n) = k_n(s_1,..,s_n)$$

et que T_2 est un intervalle fini de R, ce qui est généralement le cas puisque la durée d'observation est toujours finie.
T peut être égal à R.

Notations :

Afin d'utiliser les résultats sur les opérateurs auto-adjoints dans un Hilbert, on va écrire le modèle (1) sous une forme qui le ramène à des opérateurs linéaires.

$$x \in L^2(T_1),\ k_n \in L^2(T^n) \qquad y \in L^2(T_2)$$

En posant : $\chi_n = x(t-s_1)\ldots x(t-s_n)$, on a :

$$\int_{T^n} k_n(s_1,..,s_n)x(t-s_1)..x(t-s_n)ds_1..ds_n = \chi_n k_n$$
$$= <\chi_n, k_n>$$

la dernière expression étant le produit scalaire dans $L^2(T^n)$.

$$y = Hx = \chi k = \chi_1 k_1 + \ldots + \chi_N k_N$$

Posons : $\mathcal{L}_n = L^2(T^n)$

χ est une application de $\mathcal{L} = \mathcal{L}_1 \oplus \ldots \oplus \mathcal{L}_N \longrightarrow L^2(T_2)$

Remarquons que χ et tous les χ_n sont des transformations linéaires bornées; de plus χ est une transformation compacte (χ_n est H-S $\rightarrow$ χ_n compact $\rightarrow$ χ compact

Appellons χ^* l'opérateur adjoint de χ ; on sait que [16] [17] :

Proposition :

- $G=\chi^*\chi$ est un opérateur auto-adjoint positif compact de $\mathcal{L}$ dans $\mathcal{L}$.
- $F=\chi\,\chi^*$ est un opérateur auto-adjoint positif compact dans $L^2(T_2)$.

La première partie implique qu'il existe un ensemble de fonctions propres orthonormées $\phi_i \in \mathcal{L}$ et un ensemble de valeurs propres correspondantes $\lambda_i \geq 0$. on a donc :

$$G\phi_i = \lambda_i \phi_i \quad \text{et} \quad Gh = \Sigma\, c_i \phi_i \qquad \forall h \in \mathcal{L}$$

Pour $F=\chi\chi^*$ on montre facilement que l'ensemble des $\eta_i = \frac{1}{\sqrt{\lambda_i}} \chi\phi_i$ est un ensemble orthonormal de fonctions propres induit par (ϕ_i) dans $L^2(T_2)$:

$$F\eta_i = \lambda_i \eta_i \quad \text{et} \quad Fy = \Sigma\, d_i \eta_i \qquad y \in L^2(T_2)$$

Appellons maintenant S_x le sous espace fermé de $\mathcal{L}$ engendré par les ϕ_i correspondant aux valeurs propres non nulles.

χ étant une application : $\mathcal{L} \to L^2(T_2)$, soit χ^* son adjointe : $L^2(T_2) \to \mathcal{L}$; appellons $\mathrm{Im}(\chi^*)$ l'image de χ^*, c'est à dire l'ensemble des points de $\mathcal{L}$ obtenus par χ^*.
On a alors :

Proposition :

- L'adhérence $\overline{\mathrm{Im}(\chi^*)} = S_x$
- L'espace nul de χ (corresp. à $\lambda_i=0$) = espace nul de $G = S_x^{\perp}$

($S_x^{\perp}$ = complément orthogonal de S)

Rappellons qu'on avait écrit l'équation du système :

$$y = Hx = \chi k = \chi_1 k_1 + \ldots + \chi_N k_N \qquad (2)$$

formulation du problème :

On cherche à caractériser les classes $\mathcal{D} \subset \mathcal{L}$ déterminables par rapport à x; comme les éléments de $\mathcal{D}$ sont des H-S, il suffit de trouver leurs noyaux, et cela en fonction de l'entrée x que l'on veut déterminante ou de l'opérateurχ associé à cette entrée.

Or, déterminer $k=(k_i)$ revient à inverser l'opérateur χ dans (2).
On va decomposer k en deux éléments qu'on déterminera successivement.
On sait que :

$$\chi : \mathcal{L} \to L^2(T_2) \qquad \text{et} \qquad \mathcal{L} = S_x \oplus S_x^{\perp}$$

comme $k \in \mathcal{L}$, il admet la décomposition :

$$k = k_x + k' \qquad \text{avec } k_x \in S_x \quad \text{et} \quad k' \in S_x^{\perp}$$

Détermination des noyaux :

Proposition :

La connaissance de l'entrée x(s) et de la sortie y(t) permet de déterminer la composante k_x du noyau.
Si $b \neq 0 \rightarrow \chi b \neq 0$, alors k_x est la solution du problème (k'=0).

Preuve :

x fixé $\rightarrow$ χ fixé; on connait $G=\chi^*\chi$ et $F=\chi\chi^*$ ainsi que les systèmes (ϕ_i) et (η_i) de fonctions propres correspondantes, toutes deux associées au système (λ_i) de valeurs propres, (ϕ_i) étant un système orthonormal dans $\mathcal{L} = S_x \oplus S_x^{\perp}$ et (η_i) l'étant dans $L^2(T_2)$.

$$k_x \in S_x \rightarrow k_x = \Sigma a_i \phi_i \qquad (3)$$

k_x sera connu si on détermine les coéfficients (a_i).

Remarquons que :

$$y = \chi k = \chi k_x + \chi k' = \chi k_x$$

(car k' appartient à l'espace nul de χ d'aprés la proposition précédente)

Développons χk_x à l'aide de (3) et introduisons le système orthonormé (η_i) à l'aide de $\eta_i = \frac{1}{\sqrt{\lambda_i}} \chi\phi_i$:

$$y = \chi k_x = \chi(\Sigma a_i \phi_i) = \Sigma a_i \chi\phi_i = \Sigma a_i \sqrt{\lambda_i}\, \eta_i$$

En effectuant le produit scalaire par η_i dans la dernière expression, on obtient les coéfficients (a_i) :

$$a_i\sqrt{\lambda} = (\eta_i, y)$$

Remarque :

La solution est nécessairement incomplète, ou du moins k' ne peut être obtenu par des opérations sur y; la solution de l'équation $y=\chi k$

n'est pas unique sauf si $S_x^{\perp} = \{0\}$, auquel cas on a la solution k_x.
Autrement, on peut ajouter à la solution k_x n'importe quel élément k' de $S_x^{\perp}$ et l'équation $y=\chi k$ sera toujours satisfaite.

$$(\chi k=\chi(k_x+k')=\chi k_x+\chi k' \quad \text{et } \chi k'=0)$$

Finalement, si les valeurs propres sont $\neq 0$, la solution est unique et vaut k_x

Caractérisation des classes déterminables

x est fixé $\rightarrow S_x$ est déterminé.
On va expliciter les opérateurs déterminables par l'entrée x en définissant leurs noyaux en fonction de x ou de grandeurs liées à x.

Proposition :

Soit $\mathcal{D} \subset \mathcal{H}$ une partie bornée et déterminable par rapport à x; si C est une partie précompacte ($\bar{C}$ compacte) de S_x, alors :

$$\mathcal{D} = \{k_x + f(k_x)\} \qquad (4)$$

où $k_x \in C$, et $f(k_x)$ est une fonction continue de $\bar{C}$ dans $S_x^{\perp}$.

Réciproquement, tout ensemble de cette forme est une classe déterminable par rapport à x.

Preuve :

un tel ensemble est déterminable par rapport à x puisque pour $\delta>0$,

$$\hat{k} = \sum_{1}^{N} \frac{1}{\sqrt{\lambda_i}} (Hx,\eta_i)\phi_i \qquad \text{où } N=N(\delta)$$

satisfait :

$$\|k_x - \hat{k}\| < \delta \quad \text{pour tous les } k_x \in C$$

(Pour la déterminabilité on utilisera la norme H-S)

· Puisque f est uniformément continue, on peut choisir de telle sorte que :

$$|||f(k'_x) - f(k''_x)||| \quad \text{lorsque} \quad |||k'_x - k''_x||| < \delta \quad k'_x, k''_x \in S_x$$

alors $\hat{H} = \hat{k} + f(\hat{k})$ donne une ε-détermination.

-Réciproquement, si $\mathcal{D}$ est déterminable, sa projection C sur S_x est elle aussi déterminable et précompacte.

A chaque point k_x de C, il correspond nécessairement un seul k dans $\mathcal{D}$, puisque les mesures sur y ne peuvent déterminer que la projection de k dans S_x.
Enfin, f(k') devra être continue pour que l'estimateur $\hat{H}$ soit continu.

Corollaire :

Il découle de la définition que toute classe déterminable peut être approximée par une classe déterminable par rapport à x, c'est à dire par un ensemble de la forme (4).

Remarque :

X étant compact (X est le sous ensemble de $L^2(T_1)$ contenant les entrées x(s)), la norme du sup est plus appropriée : $\|H\| = \sup_{x \in X} \|Hx\|$

Il découle de la définition et de la comparaison des normes que si $\mathcal{D}$ est déterminable pour la norme H-S, $\mathcal{D}$ est aussi déterminable pour la norme du sup.

2.3. Difficultés numériques de la recherche d'une classe déterminable

On va expliciter les résultats précédents pour voir les difficultés d'application numérique.

Supposons que $H = H_1 + \ldots + H_N$; on déterminera la classe $\mathcal{D}_n$ pour les polynomes H_n de degré n; pour les polynomes H, il suffira de prendre :

$$\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_N\}$$

On doit déterminer toutes les grandeurs liées à x :

- D'abord χ_n :

c'est l'opérateur de $L^2(T^n) \to L^2(T_2)$ défini par le noyau $x(t-s_1)\ldots x(t-s_n)$.

$$\longrightarrow \chi_n(k_n) = \langle \chi_n, k_n \rangle = \int_{T^n} k_n(s_1,..,s_n)x(t-s_1)..x(t-s_n)ds_1..ds_n$$

$$= y(t)$$

- L'opérateur adjoint χ^* :

c'est un opérateur de même noyau allant du dual $(L^2(T_2))^* = L^2(T_2)$ dans le dual $(L^2(T^n))^* = L^2(T^n)$.

$$\longrightarrow k_n \overset{\chi_n}{\to} \chi_n k_n = y(t) \overset{\chi_n^*}{\longrightarrow} \chi_n^* y(t) = \chi_n y(t)$$

(χ_n et χ_n^* ont même noyau)

$$\longrightarrow \chi_n^* \chi_n(k_n) = \chi_n y = \int_{T_2} x(t-s_1')\ldots x(t-s_n')y(t)dt$$

$$= \int_T x(t-s_1')..x(t-s_n') \times \int_{T^n} k_n(s_1,..,s_n)x(t-s_1)..x(t-s_n)$$

$$\times ds_1 \ldots ds_n$$

- Donc $\chi_n^* \chi_n$ défini dans $L^2(T^n)$ est l'opérateur G de noyau g, c'est à dire :

$$Gk_n = \langle g, k_n \rangle$$

avec :

$$g(s_1,..,s_n;s_1',..,s_n') = \int_{T_2} x(t-s_1)..x(t-s_n)x(t-s_1')..x(t-s_n')dt$$

G est compact défini positif.

Soit $\{ \phi_i(s_1,..,s_n)\}$ un ensemble orthonormal de fonctions propres de G correspondant à un système de valeurs propres $\{ \lambda_i \}$.

Toutes les difficultés numériques résident là, dans la détermination des valeurs propres et des fonctions propres.
Si on arrive à les obtenir, on a alors le système (η_i) de fonctions propres associées à $F=\chi_n\chi_n^*$ puisqu'on sait que :

$$\eta_i = \frac{1}{\sqrt{\lambda_i}} \chi_n\phi_i \text{ avec } \chi_n\phi_i = \int_{T^n} \phi_i(s_1,.,s_n)x(t-s_1)..x(t-s_n)ds_1..ds_n$$

Posons :

$$k_n(s_1,..,s_n) = \sum_i a_i\phi_i(s_1,..,s_n) \qquad (5)$$

Les (a_i), on l'a déjà vu s'obtiennent à partir des (η_i).

L'ensemble des k_n donnés par (5) définit S_x; d'après le théorème de caractérisation, on obtient une classe $\mathcal{D}_n$ en prenant un ensemble compact de tels k_n.

P A R T I E 4

IDENTIFICATION NUMERIQUE

chap.1 : Généralités, modèle, relations entre les moments de l'entrée et de la sortie.

chap.2 : Identification des systèmes non linéaires à entrées et sorties aléatoires.

chap.3 : Identification des systèmes non linéaires
cas où les caractéristiques statistiques sont inconnues.

chap.4 : Approximation stochastique

CHAPITRE 1 : GENERALITES

On a vu que la description d'un système non linéaire pouvait être faite entièrement à partir du modèle $\hat{A}$ défini par :

$$y(t)= \hat{A}\left[x(s)/\ s \leq t\right] = \sum_{n=1}^{N} \int_{T^n} h_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n \qquad (1)$$

$\hat{A}$est entièrement déterminé dès que l'on connait les noyaux h_n.
L'identification numerique commencera au chapitre 2.

1.1. Definitions :

Posons $T=[a,b]$; le terme général du modèle est :

$$y_n(t)=H_n\left[x(s)\right] = \int_a^b..\int h_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

$$= \int..\int k_n(t,\tau_1,..,\tau_n)x(t-\tau_1)..x(t-\tau_n)d\tau_1..d\tau_n$$

$$\text{avec } k_n(t,\tau_1,..,\tau_n) = h_n(t,t-s_1,..,t-s_n)$$

$$\text{et } a \leq t \leq b$$

Les h_n ou k_n sont les noyaux; t sera dite variable d'observation et les τ_i variables d'âge.

On dira que le système est :

- realisable si $k_n(t,\tau_1,..,\tau_n)=0$ pour $\tau_i < 0$.
- a une mémoire bornée si $k_n(t,\tau_1,..,\tau_n)=0$ pour $\tau_i < \nu$ où ν est une constante positive.

Si H_n est réalisable, la borne inférieure de l'intégrale peut être prise egale à zero sans changer la valeur de l'intégrale.

1.2. Modèle pour systèmes stationnaires

Si H_n ne varie pas avec le temps (système stationnaire), s'il est réalisable et a une mémoire bornée ν, et si $(b-a)>\nu$, on peut l'écrire

$$H_n[x(t)] = \int_0^{\nu} \cdots \int k_n(\tau_1,..,\tau_n)x(t-\tau_1)..x(t-\tau_n)d\tau_1..d\tau_n$$

avec $a \leqslant t \leqslant b$

Le modèle pour systèmes stationnaires est alors :

$$y(t) = \sum_{n=1}^{N} H_n[x(t)]$$

1.3. Relations entre les moments de l'entrée et de la sortie

Toute caractérisation du processus sortie y(t) par ses moments peut être exprimée en termes de moments du processus entrée à l'aide des relations qui vont suivre.
Le modèle est pris sous sa forme générale (1); ce qui suit est valable aussi si le système n'est pas stationnaire.

Rappelons que $x \in L^2(T)$ et $h_n \in L^2(T^{n+1})$.

La fonction x(t) etant aléatoire on supposera que ses réalisations sont presque sûrement de carré sommable sur T.

On supposera aussi que :

$$E[x(t)]=0 \text{ et } E[x^2(t)] < C \qquad \forall t \quad C= \text{constante}$$

Si la condition suivante est satisfaite $\forall n$:

$$\int_{T^n} |h_n(t,s_1,..,s_n)|E|x(s_1)..x(s_n)|ds_1..ds_n < \infty \qquad (2)$$

alors la fonction :

$$y_n(t) = \int_{T^n} h_n(t,s_1,..,s_n)x(s_1)..x(s_n)ds_1..ds_n$$

existe pour chaque t et pour toute fonction x(s), sauf peut-être pour un ensemble de fonctions x(s) de probabilité nulle (Fubini).

Avec les conditions déjà imposées à x(t), il suffit pour que la condition (2) soit satisfaite que :

$$\int_{T^n} |h_n(t,s_1,..,s_n)| ds_1..ds_n < \infty \qquad \forall t$$

- L'espérance du processus sortie y(t) est :

$$Ey(t) = \sum_n \int_{T^n} h_n(t,s_1,..,s_n) E\left[x(s_1)..x(s_n)\right] ds_1..ds_n$$

- De même le moment d'ordre 2 :

$$Ey(t)y(t') = \sum_n \sum_m \int_{T^{n+m}} h_n(t,s_1,..,s_n) h_m(t',s'_1,..,s'_m) E\left[x(s_1)..x(s_n)x(s'_1)..x(s'_m)\right] ds_1..ds'_m$$

- D'une manière plus générale, toute caractérisation de la sortie y(t) par ses moments peut s'exprimer à l'aide des moments de l'entrée.
Ainsi, y_{n_i} etant le $n_i^{ème}$ terme de la série du modèle, on a :

$$Ey_{n_1}(t)y_{n_2}(t')..y_{n_p}(t^{(p-1)}) = \int_{T^{n_p}} h_{n_1}(t,s_1,..,s_n) h_{n_2}(t',s'_1,..,s'_{n_2}) \cdots$$

$$\times h_{n_p}\left[t^{(p-1)}, s_1^{(p-1)}, .., s_{n_p}^{(p-1)}\right]$$

$$\times E\left[x(s_1)..x(s_{n_1}) \ \cdots \ x(s_1^{(p-1)})..x(s_{n_p}^{(p-1)}\right.$$

$$\times ds_1 \ \cdots \ ds_{n_p}^{(p-1)}$$

CHAPITRE 2

IDENTIFICATION DES SYSTEMES NON LINEAIRES A ENTREES-SORTIES DETERMINISTES OU ALEATOIRES

Les systèmes considérés sont stationnaires (ne varient pas avec le temps).

2.1. Résolution directe

Le système a une entrée x(t) et une sortie y(t) aléatoires. (l'étude reste evidemment valable si elles sont déterministes; il suffira d'enlever le symbole espérance mathématique E dans le critère et les expressions qui en découlent).

Le problème est de déterminer l'opérateur du système à partir de l'observation de l'entrée x(t) et de la sortie y(t), ou plutôt de leurs réalisations.

On a vu que le modèle décrivant le système avait la forme :

$$y_c(t) = \hat{A}[x(s) \ / \ s \leq t] = \sum_{n=1}^{N} \int_{T^n} k_n(\tau_1,..,\tau_n)x(t-\tau_1)..x(t-\tau_n)d\tau_1..d\tau_n$$

Les noyaux k_n de ce modèle sont à déterminer de sorte que la sortie du modèle ou sortie calculée :

$$y_c = \hat{A}[x(s) \ / \ s \leq t]$$

approxime au mieux dans un sens prédéfini, la sortie vraie :

$$y(t) = A[x(s) \ / \ s \leq t]$$

(A est l'opérateur qui décrit exactement le comportement du système, et $\hat{A}$ son approximation).

On utilisera comme critère d'approximation la moyenne du carré de l'écart entre la sortie du modèle et celle du système.
On doit donc déterminer les fonctions k_n de telle sorte que, E désignant l'espérance mathématique, on ait :

$$E\left[y(t)-y_c(t)\right]^2 = \text{Minimum} \qquad (1)$$

Remarque :

Si le système est déterministe, on utilisera comme critère la minimisation de :

$$J = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N} \left[y(t_i)-y_c(t_i)\right]^2$$

t_i etant les instants de mesure.

Dans le cas général, le critère sera la minimisation de :

$$J = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \left[y(t)-y_c(t)\right]^2 dt \qquad (2)$$

Si les processus utilisés pour identifier sont stationnaires et ergodiques, on obtient un critère equivalent à (2) enremplaçant la moyenne temporelle par la moyenne statistique :

$$J = E\left[y(t)-y_c(t)\right]^2$$

La recherche du minimum de (1) peut se faire en annulant le gradient de (1) par rapport à k_n.
On obtient alors le système d'équations intégrales :

$$R_j(s_1,..,s_j) = \sum_{n=1}^{N} \int_{T^n} k_n(\tau_1,.,\tau_n)\Gamma_{j+n}(s_1,.,s_p,\tau_1,.,\tau_n)d\tau_1..d\tau_n$$

$$j=1,..,N \qquad s_i \in T$$

$$R_j(s_1,..,s_j) = E\left[y(t)x(t-s_1)..x(t-s_j)\right]$$

$$\Gamma_j(\tau_1,..,\tau_j) = E\left[x(t-s_1)...x(t-s_j)\right]$$

Ce système d'équations intégrales est complexe; sa résolution directe est en général difficile ou impossible.

Certains cas peuvent cependant être résolus directement en utilisant des processus particuliers [18] , ou en utilisant une généralisation de la méthode de factorisation spectrale [19] ; les résultats sont difficiles à mettre en oeuvre numériquement.

Citons par exemple le cas où on utilise un processus de Poisson à l'entrée; pour résoudre le système d'équations intégrales, les moments d'ordre élevé sont exprimés en termes de moments du second ordre; cela simplifie le système d'équations en le ramenant à une forme triangulaire. La dernière equation ne contient alors qu'une fonction inconnue; celle-ci est déterminée et apres substitution dans les équations précédentes, les noyaux sont obtenus successivement.

Ceci n'est pas pratique car on doit quand même passer par la résolution d'équations intégrales.

2.2. Méthode de résolution approchée

On a déjà vu que les noyaux $k_j \in L^2(T^j)$.
Supposons qu'on se fixe une base $\{\phi_i\}$ de $L^2(T)$, on obtient à partir de la base $\{\phi_i\}$, en prenant les produits tensoriels $\{\phi_{j_1} \otimes \ldots \otimes \phi_{j_n}\}$, une base de $L^2(T^n)$ (densité de $L^2(T_1) \otimes L^2(T_2)$ dans $L^2(T_1 \times T_2)$).

Les fonctions k_j cherchées s'écrivent :

$$k_j(\tau_1, .., \tau_j) = \sum h^{(j)}_{\nu_1 .. \nu_j} \phi_{\nu_1}(\tau_1) \ldots \phi_{\nu_j}(\tau_j)$$

La base (ϕ_i) etant fixée, la détermination des noyaux revient à celle des coéfficients h.

- Résolution:

La base (ϕ_i) de $L^2(T)$ comporte une infinité dénombrable de termes

Supposons que l'on prenne les M premiers termes de cette base et les N premiers termes de la série du modèle; quels sont dans ce cas les coéfficients h qui rendent minimum l'expression :

$$J = E\left[y(t)-y_c(t)\right]^2$$

En remplaçant le noyau par son expression, la sortie du modèle devient :

$$y_c(t) = \sum_{j=1}^{N} \sum_{\ell=1}^{P_j} h_\ell^{(j)} \int_{T^j} \phi_{\mu_1}(\tau_1)\ldots\phi_{\mu_j}(\tau_j)x(t-\tau_1)..x(t-\tau_j)d\tau_1..d\tau_j$$

$$P_j = \frac{(j+1)\ldots(j+M-1)}{(M-1)!}$$

$$\mu_1,\ldots,\mu_j = 1,\ldots,M$$

$$(\mu_1,\ldots,\mu_j) = \ell$$

Pour alléger l'écriture, on notera dorénavant l'espérance mathematique par une barre placée sur l'expression.

En portant l'expression de y_c dans le critère J on obtient :

$$J = \bar{y}^2 - 2 \sum_{j=1}^{N} \sum_{\ell=1}^{P_j} h^{(j)} Q_\ell^{(j)} + \sum_{j_1,j_2=1}^{N} \sum_{\ell_1,\ell_2=1}^{P_{j1}} h_{\ell_1}^{(j_1)} h_{\ell_2}^{(j_2)} B_{\ell_1\ell_2}^{(j_1,j_2)}$$

$$B_{\ell_1\ell_2}^{(j_1,j_2)} = \int_{T^{j_1+j_2}} \phi_{\mu_1}(\tau_1)..\phi_{\mu_{j_1}}(\tau_{j_1})\phi_{\nu_1}(\alpha_1)..\phi_{\nu_{j_2}}(\alpha_{j_2})$$

$$\times \overline{x(t-\tau_1)..x(t-\tau_{j_1})x(t-\alpha_1)..x(t-\alpha_{j_2})}d\tau_1\ldots d\alpha_{j_2}$$

$$Q_\ell^{(j)} = \int_{T^j} \phi_{\mu_1}(\tau_1)\ldots\phi_{\mu_j}(\tau_j)\overline{x(t-\tau_1)\ldots x(t-\tau_j)y(t)}d\tau_1..d\tau_j$$

La minimisation de J par rapport aux $h_\ell^{(j)}$ s'obtient en écrivant E(∇J)=0; cela fournit un système d'équations algébriques linéaires :

$$\underline{Bh = Q}$$

où B = matrice $B_{\ell_1\ell_2}^{(j_1,j_2)}$

Q = vecteur colonne de composantes $Q_\ell^{(j)}$

h = vecteur colonne des coéfficients inconnus $h_\ell^{(j)}$

On supposera dans ce chapitre que les caractéristiques statistique necessaires sont connues; la solution est dans ce cas :

$$h = B^{-1}Q \qquad \text{si B est inversible}$$

Souvent, en pratique la matrice B n'est pas inversible; on prend alors pour solution :

$$h = B^{+}Q \qquad \text{où } B^{+} \text{ est une matrice pseudo-inverse}^{\dagger} \text{de B}$$

† Inverse généralisée ou pseudo-inverse d'une matrice [33]

Soit à résoudre l'équation algébrique: $Ax = y$ (1)
A= m×n matrice, x= n-vecteur, y= m-vecteur.
Si A est carré et non singulière, c'est à dire si m=n=rang de A, la solution de (1) est : $x=A^{-1}y$.

Cependant, même si A^{-1} n'existe pas, il est souvent necessaire de résoudre d'une manière approchée l'equation (1), 'approchée' dans un certai sens.

Le pseudo-inverse le plus employé est celui qui utilise pour critère la minimisation par rapport à x de la forme quadratique :

$$\|Ax-y\|^2 = (Ax-y,Ax-y) \qquad (2)$$

Le x minimisant (2) peut ne pas être unique; dans ce cas on prendra comme solution le x qui a la plus petite norme et on notera : $x=A^{+}y$. A^{+} sera dite matrice pseudo-inverse ou inverse généralisée de A.

CHAPITRE 3

IDENTIFICATION DES SYSTEMES NON LINEAIRES

LES CARACTERISTIQUES STATISTIQUES SONT INCONNUES

Plus loin, (chapitre 4), on verra en détail la méthode d'approximation stochastique. On va cependant indiquer brièvement son application pour montrer l'utilité de cette méthode dans le cas où les caractéristiques statistiques sont inconnues.

Soit $Q(x,c)$ une fonction de la variable certaine c et de la variable aléatoire x, celle-ci ayant pour densité de probabilité une fonction $P(x)$.

Supposons que l'on recherche l'optimum moyen de Q, qu'on définira comme étant l'optimum de l'espérance mathématique de Q :

$$\underset{c}{\text{opt}}\ E_x\left[Q(x,c)\right]$$

L'équation déterminant la valeur c^* qui fournit l'optimum s'écrit :

$$\nabla_c E_x = E_x\left[\nabla_c\ Q(x,c)\right] = 0 \qquad (1)$$

Deux cas possibles :

- P(x) connue :

Le problème est déterministe et l'algorithme classique du gradient fournit la solution :

$$c[n] = c[n-1] + \gamma[n]\nabla_c\ (E_x Q)$$

- P(x) inconnue :

L'espérance mathématique ne peut être évaluée; on peut cependant appliquer l'algorithme ci-dessus, mais aux réalisations de la variable aléatoire représentant le gradient.

(on notera une réalisation du gradient par ∇ ou grad, et la variable aléatoire

représentant le gradient par $\hat{\nabla}$ ou $\tilde{grad}$).
L'algorithme est dans ce cas :

$$c[n] = c[n-1] + \gamma[n] \nabla_c Q(x[n], c[n-1]) \qquad (2)$$

On montrera que sous certaines conditions, cet algorithme converge avec la probabilité 1 vers la valeur c^* fournissant l'optimum.

Le problème se formule de la même manière qu'au chapitre précédent. Les équations demeurent les mêmes mais la solution $h=B^{-1}Q$ obtenue n'est plus utilisable car la matrice B fait intervenir les caractéristiques probabilistes (moments) du processus entrée x, et le vecteur Q fait intervenir celles de x et du processus sortie y. Or, ces caractéristiques sont inconnues.

On a vu que l'identification du système se ramenait à la détermination d'un vecteur h solution de l'équation $Bh=Q$.
Ecrivons cette équation :

$$Bh - Q = E[\text{grad } J] = 0$$

Ainsi écrite, cette équation est de la même forme que (1); ceci rend l'algorithme de l'approximation stochastique (2) applicable pour atteindre la solution.

Donc, les caractéristiques statistiques étant inconnues, la solution h^* est obtenue d'une manière récursive par l'algorithme :

$$h[n] = h[n-1] - \alpha[n](\text{grad}_h J[n-1]) \qquad (3)$$

$\text{grad}_h[n-1]$ étant la valeur prise à l'étape n-1 par le vecteur aléatoire $\tilde{grad}_h J(i)$, vecteur dont les composantes sont:

$$\tilde{grad}_h^{(\ell)} J[i] = \sum_{\ell} \tilde{B}_{\ell_1 \ell_2}^{(j_1, j_2)} h_{\ell}^{(j)}[i] - \tilde{Q}_{\ell}^{(j)}$$

avec :

$$\tilde{B}^{(j_1,j_2)}_{\ell_1\ell_2} = \int..\int \phi_{\mu_1}(\tau_1)..\phi_{\mu_{j_1}}(\tau_{j_1})\phi_{\nu_1}(\alpha_1)..\phi_{\nu_{j_2}}(\alpha_{j_2})$$

$$\times\, x(t-\tau_1)..x(t-\tau_{j_1})x(t-\alpha_1)..x(t-\alpha_{j_1})d\tau_1..d\tau_{j_1}d\alpha_1..d\alpha_{j_2}$$

$$\tilde{Q}^{(j)}_{\ell} = \int..\int \phi_{\mu_1}(\tau_1)..\phi_{\mu_j}(\tau_j)x(t-\tau_1)..x(t-\tau_j)y(t)d\tau_1..d\tau_j$$

$\tilde{B}$ et $\tilde{Q}$ sont des grandeurs aléatoires; l'algorithme (3) est lui-même aléatoire, mais comme on le verra au chapitre suivant, il converge avec la probabilité 1 vers la solution.

- Méthode utilisant un algorithme déterministe

Les caractéristiques statistiques sont toujours inconnues; on supposera les processus entrée et sortie stationnaires et ergodiques. Rappelons que s'il y a ergodisme, on a :

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} Q[x(t),c]dt = E_x[Q(x,c)]$$

Une méthode naturelle consiste à faire une évaluation de l'espérance en prenant la moyenne jusqu'à l'étape en cours, puis utiliser ensuite l'algorithme déterministe du gradient.
L'algorithme s'écrira dans ce cas :

$$\bar{h}[n] = \bar{h}[n-1] - \alpha[n]\frac{1}{n-1}\sum_{i=1}^{n-1} \text{grad}_{h_i} J[i] \qquad (4)$$

Dans l'algorithme (3) de l'approximation stochastique, on a $h(\infty)=h^*$, h^* étant la valeur fournissant l'optimum; on montre que l'algorithme (4) converge [21] vers la même valeur que (3) :

$$\lim_{n\to\infty} \bar{h}[n] = h[\infty] = h^*$$

Les conditions de convergence de (4) sont les même que celles de (3).

Comparaison entre les deux algorithmes

Par sa simplicité, l'algorithme (3) est plus maniable lorsqu'il s'agit de déterminer un paramètre.

Par son aspect déterministe, l'algorithme (4) est plus pratique lorsqu'il s'agit de déterminer plusieurs paramètres liés entre eux.
En effet, au lieu d'utiliser plusieurs algorithmes (un pour chaque paramètre) indépendants, on peut les lier entre eux en portant dans l'un les résultats partiels des autres à la même étape; ceci donne une approximation en bloc de tous les paramètres plus rapide et plus précise.

On peut aussi faire la même opération avec (3), mais cet algorithme , par ses oscillations, entraine un retard lors du passage entre les algorithmes des divers paramètres.

CHAPITRE 4

APPROXIMATION STOCHASTIQUE

La méthode de l'approximation stochastique peut être utilisée pour résoudre certaines équations stochastiques ou bien pour la recherche d'optimum de fonctionnelles stochastiques.

Le premier traité sur la question [20] est dû à Robbins et Monro qui recherchaient la solution d'une équation de regression.

Pour décrire la méthode on va énoncer le problème :

Problème :

à chaque valeur d'un paramètre $x \in (-\infty,+\infty)$ correspond une variable aléatoire $Y=Y(x)$ de fonction de répartition :

$$P[Y(x) < y] = F[y/x]$$

Onsuppose que la fonction de regression :

$$R(x) = E[y/x] = \int_{-\infty}^{\infty} y dF(y/x)$$

existe bien qu'inconnue, et que pour un nombre réèl donné, l'équation :

$$R(x) = \alpha \qquad (1)$$

a une solution unique θ.

Le problème est de construire une suite de variables aléatoires x_n telles que $x_n \to \theta$ lorsque $n \to \infty$.

Robbins et Monro construisirent le processus suivant :

$$x_{n+1} = x_n + a_n[\alpha - y(x_n)] \qquad (2)$$

où $y(x_n)=y_n$ est une variable aléatoire telle que :

$$P[y_n \leq y/x_n] = F(y/x_n)$$

. Ils établirent que sous certaines conditions, il y avait convergence en moyenne quadratique :

$$\lim_{n \to \infty} E\left[x_n - \theta\right]^2 = 0$$

. Avec quelques conditions en plus, on [21] qu'il y a une convergence plus forte : le processus converge avec la probabilité 1 vers θ.

Les conditions imposées à la fonction F sont fortes, mais en pratique, elles sont souvent vérifiées.

Nous allons d'abord énoncer le théorème de Robbins et Monro; on discutera ensuite des applications de la méthode et on terminera par une étude de l'optimalité de l'algorithme en recherchant son pas optimal.

Théorème :

Soit une suite réèlle a_n telle qu'il existe deux nombres c' et c" positifs, avec :

$$\frac{c'}{n} \leqslant a_n \leqslant \frac{c''}{n}$$

Supposons qu'il existe une constante c telle que :

$$P\ \{|Y(x)| \leqslant c\} = \int_{-c}^{+c} dF(y/x) = 1$$

et R(x) étant telle que :

$R(x)$ est non décroissante

$R(\theta) = \alpha$

$R'(\theta) > 0$

Alors :

$$\lim_{n \to \infty} E\left[x_n - \theta\right]^2 = 0$$

APPLICATION DE LA METHODE POUR LA RECHERCHE D'OPTIMUM DE FONCTIONS DE PLUSIEURS VARIABLES ALEATOIRES

Cas déterministe

Dans la recherche classique d'extremum d'une fonction de plusieurs variables certaines :

$$Q(c_1,\ldots,c_N) = Q(\underline{c})$$

$$\text{où} \quad \underline{c} = (c_1,\ldots,c_N)$$

La valeur $\underline{c}$ fournissant l'extremum (qu'on supposera unique) peut s'obtenir en résolvant l'équation :

$$\nabla Q(\underline{c}) = 0 \qquad \text{avec} \quad \nabla Q = \{\frac{\partial Q}{\partial c_1},\ldots,\frac{\partial Q}{\partial c_N}\}$$

Si cette équation ne peut être résolue analytiquement, sa solution peut néanmoins être atteinte par la méthode itérative du gradient dont l'algorithme est :

$$\underline{c}_n = \underline{c}_{n-1} + \gamma_n \nabla Q(\underline{c}_{n-1})$$

Le pas γ dépend de Q ; si les conditions de convergence sont satisfaites alors $\underline{c}_n \to \underline{c}$ lorsque $n \to \infty$.

Si la fonction Q n'est pas différentiable, le gradient peut être approximé par :

$$\nabla Q(\underline{c}) \sim \frac{Q_+(\underline{c},\alpha) - Q_-(\underline{c},\alpha)}{2\alpha} = J\left[\underline{c}, \alpha\right]$$

avec $Q \pm (\underline{c},\alpha) = \left\{Q(\underline{c} \pm \alpha \underline{e}_1),\ldots,Q(\underline{c} \pm \alpha \underline{e}_N)\right\}$

$\underline{e}_1 = \{1,0,\ldots,0\}$

$\underline{e}_N = \{0,\ldots,0,1\}$

et l'algorithme devient dans ce cas $\underline{c}_n = \underline{c}_{n-1} + \gamma_n J(\underline{c}_{n-1},\alpha_n)$

Cas aléatoire

La fonction $Q(\underline{x},\underline{c})$ est cette fois ci fonction de la variable certaine $\underline{c} = \{c_1,\ldots,c_N\}$ et de $\underline{x} = \{x_1,\ldots,x_N\}$ variable aléatoire ou fonction aléatoire stationnaire, dont la densité de probabilité sera notée $P(x)$.

Dans ce cas, on recherchera l'extremum de l'espérance de Q :

$$J(\underline{c}) = E_{\underline{x}}\left\{Q(\underline{x},\underline{c})\right\} = \int Q(\underline{x},\underline{c})\ P(\underline{x})d\underline{x}$$

L'équation fournissant $\underline{c}^*$ devient :

$$\nabla J(\underline{c}) = E_{\underline{x}}\left\{\nabla_c Q(\underline{x},\underline{c})\right\} = 0$$

deux cas :

- ou la densité $P(\underline{x})$ est connue, $J(\underline{c})$ peut alors être évaluée et on se ramène au cas déterministe vu précédemment.
- ou $P(x)$ n'est pas connue à priori :

  Dans ce deuxième cas l'espérance mathématique ne peut plus être évaluée ; on est alors ramené au problème de Robbins et Monro : recherche du zéro $\underline{c}^*$ de l'équation de regression $\nabla_c\left\{\int Q(\underline{x},\underline{c})\ P(\underline{x}/\underline{c})d\underline{x}\right\} = 0$. On peut alors appliquer la méthode de l'approximation stochastique ; cette méthode s'avère très intéressante en pratique car elle revient à appliquer l'algorithme classique du gradient aux réalisations de la fonction $\nabla_c Q(\underline{x},\underline{c})$. On a donc l'algorithme :

$$\underline{c}[n] = \underline{c}[n-1] + \gamma[n]\ \nabla_c\ Q(\underline{x}[n],\ \underline{c}[n-1]) \qquad (1)$$

(on notera indifféremment c_n ou $c[n]$ la $n^{ème}$ itération).

Cet algorithme est aléatoire puisqu'il s'applique à des réalisations d'une quantité aléatoire ; sa convergence devra donc être considérée dans un sens probabiliste.

Conditions de convergence

L'algorithme (1) converge en moyenne quadratique :

$$\lim_{n\to\infty} E\left\{ \left\| \underline{c}[n] - \underline{c}^* \right\| \right\}^2 = 0$$

($\| \quad \|$ parce que $\underline{c}$ est un vecteur).

il converge même avec la probabilité 1 : $P\left\{ \lim_{n\to\infty} (\underline{c}[n] - \underline{c}^*) = 0 \right\} = 1$ lorsque les conditions suivantes sont satisfaites : (τ désignera transposition).

- $\sum_{n=1}^{\infty} \gamma[n] = \infty \qquad \sum_{n=1}^{\infty} \gamma^2[n] < \infty$

- $\inf E_x \left\{ (\underline{c} - \underline{c}^*)^\tau \nabla_c Q(\underline{x},\underline{c}) \right\} \leqslant 0 \qquad \varepsilon > 0$

  $\varepsilon \leqslant \| \underline{c} - \underline{c}^* \| \leqslant \frac{1}{\varepsilon}$

- $E_x \left\{ \nabla_c^\tau Q(\underline{x},\underline{c}) \nabla_c Q(\underline{x},\underline{c}) \right\} \leqslant d\ (1+\underline{c}^\tau \underline{c}) \qquad d > 0$

Ces conditions peuvent être interprétées comme suit :

- la première signifie que la vitesse à laquelle $\gamma[n]$ décroit doit être suffisamment rapide pour que la variance d'une estimation de $J(\underline{c})$ puisse s'annuler, mais doit aussi être suffisamment lente pour fournir assez de données permettant l'application de la loi des grands nombres.
- la deuxième détermine le comportement de la surface $E_x\{\nabla_c(\underline{x},\underline{c})\} = 0$ au voisinage de la racine $\underline{c}^*$ et en conséquence détermine les signes des increments dans $\underline{c}[n]$.
- la troisième établit que la moyenne $E_x\{\nabla_c^\tau Q(\underline{x},\underline{c})\ \nabla_c Q(\underline{x},\underline{c})\}$ ne doit pas augmenter avec $\underline{c}$ plus rapidement qu'une certaine fonction quadratique.

Version continue de la méthode

Les algorithmes donnés peuvent être considérés comme des équations aux différences non linéaires ; la convergence des algorithmes correspondrait alors à la stabilité des solutions des équations aux différences correspondantes. Pour les algorithmes déterministes la convergence correspond à la stabilité ordinaire et dans le cas stochastique la convergence est associée à la stabilité au sens stochastique, c'est-à-dire en probabilité ou avec la probabilité 1.

Par passage à la limite de l'équation aux différences à l'équation différentielle, on associe aux deux algorithmes (déterministe et stochastique) les algorithmes continus :

$$\frac{d\underline{c}(t)}{dt} = \gamma(t)\nabla Q(\underline{c}(t)) \qquad \text{déterministe}$$

$$\frac{d\underline{c}(t)}{dt} = \gamma(t)\nabla_{\underline{c}} Q(\underline{x}(t), \underline{c}(t)) \qquad \text{stochastique}$$

Les conditions de convergence dans ce cas seraient :

$$\gamma(t) > 0 \qquad \int_1^\infty \gamma(t)dt = \infty \qquad \int_1^\infty \gamma^2(t)dt < \infty$$

La résolution de l'équation différentielle fournit $\underline{c}^*$.

Exemple

Estimation d'une fonction densité de probabilité P(x) à partir de points observés x_k ($k = 1, \ldots, n$).

On supposera $P(x) \in L^2$, on se fixe une base de fonctions orthonormales $\phi_\nu(x)$ et on cherche les coefficients c_ν pour que l'écart quadratique moyen entre l'estimation $\hat{P}(x)$ de P(x) par un développement fini $\sum_{\nu=1}^{N} c_\nu\phi_\nu(x)$, et P(x) soit minimum. Nous devons donc résoudre :

$$\min_{c_\nu} \int_X \left[P(x) - \sum_{\nu=1}^{N} c_\nu\phi_\nu(x)\right]^2 dx \qquad (1)$$

En différentiant (1) par rapport aux c_ν, en égalant les dérivées à zéro et d'après l'orthonormalité des ϕ_ν

$$(\int \phi_\nu(x)\phi_\mu(x)dx = \begin{cases} 0 & \mu \neq \nu \\ 1 & \mu = \nu \end{cases}$$

On obtient :

$$\int_X P(x)\ \phi_\nu(x)dx = c_\nu \qquad \nu = 1,\ldots,N$$

P(x) étant la densité de probabilité, les coefficients c_ν de l'estimation optimale $\hat{P}(x)$ sont égaux aux espérances mathématiques des fonctions $\phi_\nu(x)$:

$$c_\nu^* = E\left[\phi_\nu(x)\right] \qquad \nu = 1,\ldots,N \qquad (2)$$

Pour appliquer l'approximation stochastique la relation (2) peut s'écrire

$$E\{\phi_\nu(x) - c_\nu\} = 0 \qquad \nu = 1,\ldots,N$$

L'algorithme de Robbins Monro donne :

$$c_\nu\left[n+1\right] = c_\nu[n] + \gamma\left[n\right]\left\{\phi_\nu(x[n]\ ) - c_\nu[n]\right\}$$
$$\nu = 1,\ldots,N$$

Et dans ce cas $c_\nu[n] \longrightarrow c_\nu^*$ avec la probabilité 1 lorsque $n \to \infty$ si on a :

$$\gamma_n > 0 \qquad \sum_{n=1}^{\infty} \gamma_n = \infty \qquad\qquad \sum_{n=1}^{\infty} \gamma_n^2 < \infty$$

OPTIMALITE DE L'ALGORITHME DU POINT DE VUE CHOIX DU PAS OPTIMAL

Cas de l'exemple précédent

Ce sera en général le cas où le gradient est linéaire en c ; c'était le cas dans l'exemple précédent lorsqu'on recherchait le minimum de

$$J(c) = E_x \left\{ [c - \phi(x)]^2 \right\} \qquad (1)$$

Le minimum était atteint pour la valeur $c^* = E_x\{\phi(x)\}$

En écrivant cela sous forme d'équation de regression : $E\{c-\phi(x)\}=C$ on obtenait par application de Robbins Monro l'algorithme :

$$c[n] = c[n-1] + \gamma[n]\,(\phi(x[n]) - c[n-1]) \qquad (2)$$

qui convergeait presque sûrement si

$$\gamma(n) > 0 \qquad \sum_{n=1}^{\infty} \gamma(n) = \infty \qquad \sum_{n=1}^{\infty} \gamma^2(n) < \infty$$

Les propriétés de l'algorithme dépendent du choix du pas $\gamma[n]$.

γ devra d'abord satisfaire les conditions (3). Le choix sera ensuite fait d'après le critère suivant :

l'algorithme (2) sera dit optimal (du point de vue du pas) si l'écart quadratique moyen :

$$E\{(c[n] - c^*)^2\} = V^2[n] \qquad (4)$$

est minimal à chaque pas.

La substitution de l'algorithme (2) dans (4) donne :

$$V^2[n] = E\left\{\left[\eta[n-1] - \gamma[n]\,(\eta[n-1] + \xi[n])\right]^2\right\} \qquad (5)$$

avec
$$\eta[n] = c[n] - c^*$$
$$\xi[n] = c^* - \phi(x[n])$$
$$E\left\{\xi^2[n]\right\} = \sigma^2$$

on peut écrire aussi :

$$V^2[n] = (1-\gamma[n])^2\ V^2[n-1] + \gamma^2[n]\sigma^2$$

Le minimum de cette expression est atteint par :

$$\gamma_{opt}[n] = \frac{V^2[n-1]}{V^2[n-1] + \sigma^2} \qquad (6)$$

Pour cette valeur, $V^2[n]$ est minimal et a pour valeur :

$$V^2[n] = \frac{V^2[n-1]\ \sigma^2}{V^2[n-1] + \sigma^2} \qquad (7)$$

de (6) et (7) on tire :

$$\gamma_{opt}[n] = \frac{1}{n + \dfrac{\sigma^2}{V^2[0]}} \qquad V^2[n] = \frac{\sigma^2}{n + \dfrac{\sigma^2}{V^2[0]}} \qquad (8)$$

L'algorithme est donc optimal au sens défini si on prend pour pas l'expression γ_{opt} de (8). Les conditions de convergences sont satisfaites pour γ_{opt}.

Remarque : S'il n'y a aucune information à priori sur la dispersion initiale, $V^2[0]$ est pris égal à l'infini $\gamma_{opt}[n] = \frac{1}{n}$ et

$$V^2[n] = \frac{\sigma^2}{n}$$

Cas général

Le problème est de minimiser l'expression

$$J(c) = E\{Q(x,c)\} \qquad (1)$$

où Q est une fonction différentiable en c.

La valeur c* du paramètre qui minimise (1) est donnée par l'algorithme :

$$c[n] = c[n-1] - \gamma[n] \nabla_c Q(x[n], c[n-1])$$

on supposera qu'il y a ergodisme pour les fonctions aléatoires x.

Etant donné l'ergodisme, on va raisonner sur la moyenne de l'échantillon :

$$J[n] = \frac{1}{n} \sum_{i=1}^{n} Q(x[i], c[n]) \qquad (2)$$

on va déterminer $\gamma[n]$ de telle sorte que (2) soit minimale pour chaque valeur de n.

Si on note par Q'_c la dérivée par rapport à c, la condition de minimum s'écrit :

$$\sum_{i=1}^{n} Q'_c(x[i], c[n]) = 0 \qquad (3)$$

ou bien

$$\sum_{i=1}^{n} Q'_c(x[i], c[n-1] - \gamma[n] Q'_c(x[n], c[n-1])) = 0$$

L'équation (3) n'est pas linéaire en $\gamma[n]$; On va la résoudre de façon approchée en supposant $\gamma[n] \ll 1$ et en la linéarisant. L'algorithme que l'on obtiendra dans ce cas sera dit quasi-optimal.

En faisant un développement limité suivant les puissances de $\gamma[n]$, de l'équation (3), et en se limitant à l'approximation linéaire on obtient :

$$\sum_{i=1}^{n} Q'_c(x[i], c[n-1]) - \gamma[n] Q'_c(x[n], c[n-1]) \sum_{i=1}^{n} Q''_c(x[i], c[n-1]) = 0$$

jusqu'à l'étape n-1, l'optimisation a eu lieu, c'est-à-dire que tous les pas ont été choisis optimalement, donc d'après les expressions (2) et (3) on a $\nabla_c J[n-1] = 0$ ou bien :

$$\sum_{i=1}^{n-1} Q'_c(x[i], c[n-1]) = 0$$

d'où le pas "quasi-optimal" :

$$\gamma_{qopt}[n] = \frac{1}{\sum_{i=1}^{n} Q''_c(x[i], c[n-1])}$$

Ceci a lieu lorsque $\gamma << 1$ ou $|c[n] - c[n-1]| << 1$.

Les conditions de convergence sont satisfaites pour γ_{qopt}.

P A R T I E 5

SIMULATION

1. Calcul formel d'identification
   Algorithmes pratiques pour l'approximation par un opérateur H-S d'ordre 2.

2. Application numérique; exemple.

IDENTIFICATION D'UN SYSTEME NON LINEAIRE PAR UN H-S D'ORDRE 2 ALGORITHMES PRATIQUES ET EXEMPLE NUMERIQUE

On va décrire la détermination d'un système non linéaire en l'approximant par un H-S d'ordre 2.

1. Calcul formel d'identification

Hypothèses

Le système est approximé par un polynome H-S d'ordre 2; (ϕ_i) étant une base de $L^2(T)$, la sortie du système (ou plutôt du modèle qui le représente) y(t) correspondant à l'entrée x(z) est donnée par :

$$y(t) = \int_T k_1(t-z)x(z)dz + \int_T \int_T k_2(t-z_1, t-z_2)x(z_1)x(z_2)dz_1dz_2$$

- $k_1(t) = \sum_{i=1}^{N} \alpha_i \phi_i(t)$

- $k_2(t_1,t_2) = \sum_{i,j=1}^{N} \beta_{ij} \phi_i(t_1)\phi_j(t_2)$

Données :

entrées mesurées : vecteur $x(t_k)$ $k=1,..,P$
sorties mesurées : vecteur $y'(t_k)$ $k=1,..,P$

P couples $[x(t_k), y'(t_k)]$

Problème :

Trouver les N coéfficients α_i et les N^2 coéfficients β_{ij}.

CALCULS :

$$y(t_k) = y_1(t_k) + y_2(t_k)$$

avec
$$\begin{cases} y_1(t_k) = \int_T k_1(t_k - \zeta)\, x(\zeta)\, d\zeta \\ y_2(t_k) = \int_T \int_T k_2(t_k - \zeta_1, t_k - \zeta_2)\, x(\zeta_1)\, x(\zeta_2)\, d\zeta_1 d\zeta_2 \end{cases}$$

Expression de $y_1(t_k)$

$$y_1(t_k) = \int_T \left[\sum_{i=1}^{N} \alpha_i \phi_i(t_k - \zeta) \right] x(\zeta) d\zeta$$
$$= \sum_{i=1}^{N} \int_T \phi_i(t_k - \zeta)\, x(\zeta)\, d\zeta$$

On applique la méthode des trapèzes pour calculer $\int_T \phi_i(t_k - \zeta) x(\zeta) d\zeta$
Pour cela, on divise l'intervalle $[0,T]$ en D sous intervalles d'amplitude $\frac{T}{D} =$
cette intégrale devient :

$$\int_T \phi_i(t_k - \zeta) x(\zeta) d\zeta = h \sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell)\, x(\zeta_\ell)$$

avec $\zeta_\ell \in [(\ell-1)h, \ell h]$

donc
$$y_1(t_k) = \sum_{i=1}^{N} h \alpha_i \left[\sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell)\, x(\zeta_\ell) \right] \qquad (1)$$

Expression de $y_2(t_k)$

$$y_2(t_k) = \int_T \int_T k_2(t_k - \zeta_1, t_k - \zeta_2)\, x(\zeta_1)\, x(\zeta_2)\, d\zeta_1 d\zeta_2$$

$$= \int_T \int_T \left\{ \sum_{i,j=1}^{N} \beta_{ij} \phi_i(t_k - \zeta_1) \phi_j(t_k - \zeta_2) \right\} x(\zeta_1)\, x(\zeta_2)\, d\zeta_1 d\zeta_2$$

$$= \sum_{i,j=1}^{N} \beta_{ij} \left\{ \int_T \int_T \phi_i(t_k - \zeta_1)\, \phi_j(t_k - \zeta_2)\, x(\zeta_1)\, x(\zeta_2)\, d\zeta_1 d\zeta_2 \right\}$$

$$= \sum_{i,j=1}^{N} \beta_{ij} \left\{ \left(\int_T \phi_i(t_k - \zeta_1) x(\zeta_1) d\zeta_1 \right) \left(\int_T \phi_j(t_k - \zeta_2) x(\zeta_2) d\zeta_2 \right) \right\}$$

En appliquant (comme pour le calcul de y_1) la méthode des trapèzes, on obtient :

$$\int_T \phi_i(t_k - \zeta_1)\, x(\zeta_1)\, d\zeta_1 = h \sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell)\, x(\zeta_\ell)$$

$$\int_T \phi_j(t_k - \zeta_2)\, x(\zeta_2)\, d\zeta_2 = h \sum_{\ell=1}^{D} \phi_j(t_k - \zeta_\ell)\, x(\zeta_\ell)$$

donc :

$$y_2(t_k) = \sum_{i,j=1}^{N} h^2 \beta_{ij} \left\{ \left(\sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell)\, x(\zeta_\ell) \right) \left(\sum_{\ell=1}^{D} \phi_j(t_k - \zeta_\ell)\, x(\zeta_\ell) \right) \right\} \quad (2$$

Expression de $y(t_k)$

$$y(t_k) = y_1(t_k) + y_2(t_k)$$

En remplaçant $y_1(t_k)$ par son expression (1) et $y_2(t_k)$ par (2), on obtient :

$$y(t_k) = \sum_{i=1}^{N} h\alpha_i \left\{ \sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell)\, x(\zeta_\ell) \right\}$$

$$+ \sum_{i,j=1}^{N} h^2 \beta_{ij} \left\{ \left(\sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell)\, x(\zeta_\ell) \right) . \left(\sum_{\ell=1}^{D} \phi_j(t_k - \zeta_\ell)\, x(\zeta_\ell) \right) \right.$$

$$y(t_k) = h \left\{ \sum_{i=1}^{N} \alpha_i \left(\sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell) \right) \right\} \quad (3)$$

$$+ h^2 \left\{ \sum_{i,j=1}^{N} \beta_{ij} \left[\left(\sum_{\ell=1}^{D} \phi_i(t_k - \zeta_\ell) \right) . \left(\sum_{\ell=1}^{D} \phi_j(t_k - \zeta_\ell)\, x(\zeta_\ell) \right) \right] \right\}$$

Remarque :

$$\phi_i(u)\, \phi_j(u) = \phi_j(u)\, \phi_i(u) \quad \forall u$$

On peut donc écrire :

$$k_2(u_1, u_2) = \sum_{i \neq j = 1}^{N} 2\, \beta_{ij} \phi_i(u_1) \phi_j(u_2) + \sum_{i=1}^{N} \beta_{ii} \phi_i(u_1) \phi_i(u_2)$$

L'expression de $y_2(t_k)$ devient :

$$y_2(t_k) = h^2 \sum_{i \neq j=1}^{N} 2\beta_{ij} \left[\left(\sum_{\ell=1}^{D} \phi_i(t_k - \gamma_\ell) x(\gamma_\ell) \right) \left[\sum_{\ell=1}^{D} \phi_j(t_k - \gamma_\ell) x(\gamma_\ell) \right] \right] + h^2 \sum_{i=1}^{N} \beta_{ii} \left[\sum_{\ell=1}^{D} \phi_i(t_k - \gamma_\ell) x(\gamma_\ell) \right]^2$$

et l'expression de $y(t_k)$ devient :

$$\left[\begin{array}{l} y(t_k) = h \left\{ \sum_{i=1}^{N} \alpha_i \left[\sum_{\ell=1}^{D} \phi_i(t_k - \gamma_\ell) x(\gamma_\ell) \right] \right\} \\ \quad + 2h^2 \left\{ \sum_{i \neq j=1}^{N} \beta_{ij} \left[\sum_{\ell=1}^{D} \phi_i(t_k - \gamma_\ell) x(\gamma_\ell) \right] \cdot \left[\sum_{\ell=1}^{D} \phi_j(t_k - \gamma_\ell) x(\gamma_\ell) \right] \right\} \\ \quad + h^2 \left\{ \sum_{i=1}^{N} \beta_{ii} \left[\sum_{\ell=1}^{D} \phi_i(t_k - \gamma_\ell) x(\gamma_\ell) \right]^2 \right\} \end{array} \right. \quad (5)$$

Nombre d'inconnues :

$N : \alpha_i$

$N : \beta_{ii}$

$\frac{N^2 - N}{2} : \beta_{ij}$

Moindres carrés :

y et y' étant respectivement les sorties calculée et mesurée :

$$S_p = (y(t_p) - y'(t_p))$$

$$S = \sum_{p=1}^{P} (y(t_k) - y'(t_p))^2 = \sum_{p=1}^{P} (S_p)^2$$

$$\frac{1}{2} \frac{\partial S}{\partial \alpha_i} = \sum_{p=1}^{P} S_p \frac{\partial S_p}{\partial \alpha_i} = 0 \qquad i = 1, \ldots, N$$

$$\frac{1}{2} \frac{\partial S}{\partial \beta_{ii}} = \sum_{p=1}^{P} S_p \frac{\partial S_p}{\partial \beta_{ii}} = 0 \qquad i = 1, \ldots, N$$

$$\frac{1}{2} \frac{\partial S}{\partial \beta_{ij}} = \sum_{p=1}^{P} S_p \frac{\partial S_p}{\partial \beta_{ij}} = 0 \qquad i,j \; i \neq j \; 1, \ldots, N$$

On a un système de $\left(N + N + \frac{N^2 - N}{2}\right)$ équations.

On pose : $F(n_1, n_2, n_3) = \phi_{n_1}(t_{n_2} - z_{n_3})\, x(z_{n_3})$

$$\frac{\partial S_p}{\partial \alpha_i} = h\left[\sum_{l=1}^{D} F(i,p,l)\right]$$

$$\frac{\partial S_p}{\partial \beta_{ii}} = h^2\left[\sum_{l=1}^{D} F(i,p,l)\right]^2$$

$$\frac{\partial S_p}{\partial \beta_{ij}} = 2h^2\left[\sum_{l=1}^{D} F(i,p,l)\right].\left[\sum_{l=1}^{D} F(j,p,l)\right]$$

$$S_p = h\left\{\sum_{i=1}^{N} \alpha_i\left[\sum_{l=1}^{D} F(i,p,l)\right]\right\}$$

$$+ h^2\left\{\sum_{i=1}^{N} \beta_{ii}\left[\sum_{l=1}^{D} F(i,p,l)\right]^2\right\}$$

$$+2h^2\left\{\sum_{\substack{1\\ i\neq j}}^{N} \beta_{ij}\left[\left(\sum_{l=1}^{D} F(i,p,l)\right).\left(\sum_{l=1}^{D} F(j,p,l)\right)\right]\right\} - y'(t_p)$$

Formation du vecteur V des inconnues de dimension $2N + \frac{N^2 - N}{2}$

$\alpha_i = V(i)$ avec $i = 1, \ldots, N$

$\beta_{ii} = V(N+i)$ " $i = 1, \ldots, N$

$\beta_{ij} = V(2N+k)$ $i = 1, \ldots, N-1$ et $j > i$ $\quad k = (i-1)\left(N - \frac{i}{2}\right) + j - i$

On pose

$$G(i,j) = \sum_{l=1}^{D} F(i,j,l) \text{ avec } \quad F(i,j,l) = \phi_i(t_j - z_l)\, x(z_l)$$

Donc :

$$S_p = h\sum_{i=1}^{N} \alpha_i\, G(i,p) + h^2 \sum_{i=1}^{N} \beta_{ii}\,(G(i,p))^2$$

$$+ 2h^2 \sum_{\substack{i=1\\ j>i}}^{N-1} \beta_{ij}\, G(i,p).G(j,p) - y'(t_p)$$

$$\frac{1}{2}\frac{\partial S}{\partial \alpha_i} = \sum_{p=1}^{P} S_p \frac{\partial S_p}{\partial \alpha_i} = 0 \qquad \text{avec} \qquad \frac{\partial S_p}{\partial \alpha_i} = h\, G(i,p)$$

$$\frac{1}{2}\frac{\partial S}{\partial \beta_{ii}} = \sum_{p=1}^{P} S_p \frac{\partial S_p}{\partial \beta_{ii}} = 0 \qquad \frac{\partial S_p}{\partial \beta_{ii}} = h^2 [G(i,p)]^2$$

$$\frac{1}{2}\frac{\partial S}{\partial \beta_{ij}} = \sum_{p=1}^{P} S_p \frac{\partial S_p}{\partial \beta_{ij}} = 0 \qquad \frac{\partial S_p}{\partial \beta_{ij}} = 2 h^2 G(i,p)\, G(j,p)$$

Coéfficients de la matrice

Cas 1 : $r \leq N$
$s \leq N$

$$B(r,s) = \sum_{p=1}^{P} \left\{ h\, G(r,p)\, h\, G(s,p) \right\}$$

$$H(r) = \sum_{p=1}^{P} \left\{ h\, G(r,p)\, y'(t_p) \right\}$$

Cas 2 :

$r \leq N$

$N < s \leq 2N \quad s - N = i$

$$B(r,s) = \sum_{p=1}^{P} \left\{ h\, G(r,p) \cdot h^2 [G(i,p)]^2 \right\}$$

Cas 3 :

$r \leq N$

$2N < s \leq 2N + \frac{N^2 - N}{2}$

$$s = 2N + (i-1)\left(N - \frac{i}{2}\right) + j - i$$

$$i = 1, \ldots, N-1 \qquad j > i$$

$$B(r,s) = \sum_{p=1}^{P} \left\{ h\, G(r,p) \cdot 2h^2 \cdot G(i,p) \cdot G(j,p) \right\}$$

Cas 1' :

$$N < r \leq 2N$$

$$1 \leq s \leq N \qquad i = r - N$$

$$B(r,s) = \sum_{p=1}^{P} \left\{ h^2 \left[G(i,p) \right]^2 \cdot h\, G(s,p) \right\}$$

$$H(r) = \sum_{p=1}^{P} \left\{ h^2 \left[G(i,p) \right]^2 \cdot y'(t_p) \right\}$$

cas 2' :

$$N < r \leq 2N \qquad i = r - N$$

$$N < s \leq 2N \qquad k = s - N$$

$$B(r,s) = \sum_{p=1}^{P} \left\{ h^2 \left[G(i,p) \right]^2 \cdot h^2 \left[G(k,p) \right]^2 \right\}$$

cas 3' :

$$N < r \leq 2N \qquad i = r - N$$

$$2N < s \leq 2N + \frac{N^2 - N}{2} \qquad s = 2N + (u-1)\cdot\left(N - \frac{u}{2}\right) + v - u$$

$$u = 1, \ldots, N-1$$

$$v > u$$

$$B(r,s) = \sum_{p=1}^{D} \left\{ h^2 \left[G(i,p) \right]^2 + 2h^2\, G(u,p) \cdot G(v,p) \right\}$$

Cas 1''

$2N < r \leq 2N + \frac{N^2-N}{2}$ $\qquad r = 2N + (u-1)\left(N - \frac{u}{2}\right) + v - u$

$s \leq N$

$$B(r,s) = \sum_{p=1}^{P} \left\{ 2h^2\, G(u,p).\, G(v,p).\, h\, G(s,p) \right\}$$

$$H(r) = \sum_{p=1}^{P} \left\{ 2h^2\, G(u,p).\, G(v,p)\, y''(t_p) \right\}$$

Cas 2''

$2N < r \leq 2N - \frac{N^2-N}{2}$ $\qquad r = f(u,v)$

$N < s \leq 2N$ $\qquad i = s - N$

$$B(r,s) = \sum_{p=1}^{P} \left\{ 2h^2\, G(u,p).\, G(v,p).\, h^2 [G(i,p)]^2 \right\}$$

Cas 3''

$2N < r \leq 2N - \frac{N^2-N}{2}$ $\longrightarrow$ $r = 2N + (u-1)\left(N - \frac{u}{2}\right) + v - u$

$2N < s \leq 2N - \frac{N^2-N}{2}$ $\qquad s = 2N + (i-1)\left(N - \frac{i}{2}\right) + j - i$

$$B(r,s) = \sum_{p=1}^{P} \left\{ 2h^2\, G(u,p).\, G(v,p).\, 2h^2\, G(i,p)\, G(j,p) \right\}$$

2. Application numérique

Le schéma précédent est appliqué à l'identification du système non linéaire A décrit par l'opérateur suivant :

L'entrée x(t) est discrète et prend les valeurs :

$$x(t_1), x(t_2), \ldots, x(t_n), \ldots$$

La sortie y(t) est définie aux instants t_j par :

$$y(t_1) = [x(t_1)]^2$$

$$y(t_2) = [x(t_1)]^2 + [x(t_2)]^2$$

. . . . . .

$$y(t_i) = \sum_{j=1}^{i} [x(t_j)]^2$$

. . . . . .

La correspondance entre l'entrée et la sortie est non linéaire et fait intervenir le passé (la sortie à l'instant t est fonction de toutes les valeurs de l'entrée x(s), s variant de l'instant initial jusqu'à l'instant t).

L'application de l'étude précédente fournit une approximation $\hat{A}$ de A, et cela à partir de mesures sur une entrée déterminante x(t).

Test de l'identification obtenue

Une fois l'approximation $\hat{A}$ de l'opérateur obtenue et fixée à partir d'une entrée déterminante x(t), on connait deux expressions de l'opérateur :

l'une exacte A fournie par le schéma même de l'exemple : $(Ax)(t) = \sum_{1}^{i} [x(t_j)]^2$ et l'autre approchée $\hat{A}$ qui est le modèle H-S d'ordre 2.

On peut alors utiliser une entrée test de vérification $x_v(t)$, pour tester l'approximation obtenue.

A partir de cette entrée test, on détermine les deux groupes de sorties :

- Les sorties exactes $y(t_1),\ldots,y(t_n)$ aux instants $t_1,..,t_n$, cela par application de la forme exacte de l'opérateur.
- Les sorties calculées ou sorties du modèle : $y_c(t_1),\ldots,y_c(t_n)$

(Rappelons que ce deuxieme groupe de sorties est calculé par application du modèle H-S, dans l'ignorance totale de l'opérateur exact A du système)

Les résultats $\{y(t_1),..,y(t_n)\}$ et $\{y_c(t_1),..,y_c(t_n)\}$ sorties du système et du modèle, correspondant à l'entrée test seront exposés et comparés.

Remarque :

Si on utilise une seule entrée x(t) pour déterminer le modèle $\hat{A}$, on obtient en général :

$$A(x) \simeq \hat{A}(x)$$

(modèle et système ont des résultats voisins pour x)
mais A(x') est totalement différent de $\hat{A}(x')$ si $x \neq x'$.

Notre identification sera faite à partir de plusieurs fonctions entrées.

Les résultats seront exposés comme suit :

- Le temps est discrétisé aux instants $t_1, t_2, .., t_n, ..$

  $t_1 = T(1)$, $t_2 = T(2), .., t_j = T(J), ..$

- On note X(J) l'entrée pour déterminer l'opérateur.
- $Y_M(J)$ est la sortie exacte (ou mesurée) de l'opérateur, à l'instant T(J)
- SC1(J) = sortie calculée (ou sortie du modèle) à l'instant T(J).
  (l'indice 1 indique qu'on travaille selon la base 1 du programme , qui est la base des fonctions en escaliers)
- XV(J) = entrée test pour vérifier que l'approximation obtenue est valable pour prédire la sortie correspondant à une entrée autre que celle utilisée pour déterminer l'opérateur.
- YVM(J) = sortie exacte (ou mesurée) correspondant à l'entrée test.
- ZC1(J) = sortie calculée (ou sortie du modèle) correspondant à l'entrée test.

Le tableau des résultats numériques est présenté comme suit :

J	T(J)	X(J)	YM(J)	SC1(J)		T(J)	XV(J)	YVM(J)	ZC1(J)
<----identification---->						<---- controle ou test de l'identification ---->			

Details numériques

L'intervalle $(t_1, .., t_n)$ est pris égal à (0,1); il est divisé en P points et dans chaque intervalle, on utilise N fonctions de base pour calculer les intégrales intervenant dans le modèle.

Les résultats imprimés correspondent à P=70 N=5 .

La base 1 est celle des fonctions en escalier. Le programme laisse le choix de P et N libre et offre la possibilité de travailler sur trois bases àla fois (la base 2 est écrite avec les 2 premiers éléments de la base de Laguerre).

L'entrée utilisée pour la détermination des noyaux de l'opérateur est définie comme suit :

l'intervalle $[0,1]$ est divisé en M sous intervalles, chacun de ces sous intervalles est divisé en dix parties.

La fonction entrée sera égale successivement à : $1+t$, t^2, e^t, e^{-t}, sint, cost, $1+t$, t, t^3+2, t^3 dans chacune des parties du sous intervalle. Elle se répete de la même manière au prochain sous intervalle, et ainsi de suite jusqu'à épuisement.
(cela revient à simuler un processus à l'entrée).

L'entrée de test correspondant aux résultats imprimés est :

$$x_v(t) = 1 + t + t^2 + t^3$$

IBASE1 = 1

N2 = 2
IBASE2 = 0

NOMBRE DE POINTS = 70

J = NUMERO DE LA MESURE
T(J) = INSTANT DE LA MESURE NUMERO J
X(J) = ENTREE DETERMINANTE A INSTANT T(J)
YM(J) = SORTIE DETERMINANTE MESUREE A INSTANT T(J)
SC1(J) = SORTIE CALCULEE CORRESPONDANTE A ENTREE DETERMINANTE X(J) POUR BASE1
SC2(J) = SORTIE CALCULEE CORRESPONDANTE A ENTREE DETERMINANTE X(J) POUR BASE2

XV(J) = ENTREE POUR TESTER OPERATEUR DETERMINE PAR X(J)
YVM(J) = SORTIE DE CONTROLE CORRESPONDANT A ENTREE TEST XV(J)
ZC1(J) = SORTIE CORRESPONDANT A ENTREE TEST CALCULEE PAR OPERATEUR BASE1
ZC2(J) = SORTIE CORRESPONDANT A ENTREE TEST CALCULEE PAR OPERATEUR BASE2

J	T(J)	X(J)	YM(J)	SC1(J)	SC2(J)	TV(J)	XV(J)	YVM(J)	ZC1(J)	ZC2(J)
1	0.0	0.100E+01	0.100E+01	0.107E+01	0.0	0.0	0.100E+01	0.100E+01	0.107E+01	0.0
2	0.014	0.210E-03	0.100E+01	0.107E+01	0.0	0.145E-01	0.101E+01	0.203E+01	0.223E+01	0.0
3	0.029	0.244E-04	0.100E+01	0.107E+01	0.0	0.290E-01	0.103E+01	0.309E+01	0.347E+01	0.0
4	0.043	0.200E+01	0.500E+01	0.341E+01	0.0	0.435E-01	0.105E+01	0.418E+01	0.481E+01	0.0
5	0.058	0.105E+01	0.612E+01	0.477E+01	0.0	0.580E-01	0.106E+01	0.531E+01	0.624E+01	0.0
6	0.072	0.930E+00	0.699E+01	0.601E+01	0.0	0.725E-01	0.108E+01	0.647E+01	0.777E+01	0.0
7	0.087	0.868E-01	0.700E+01	0.613E+01	0.0	0.870E-01	0.110E+01	0.767E+01	0.940E+01	0.0
8	0.101	0.995E+00	0.799E+01	0.754E+01	0.0	0.101E+00	0.111E+01	0.891E+01	0.112E+02	0.0
9	0.116	0.115E+00	0.800E+01	0.771E+01	0.0	0.116E+00	0.113E+01	0.102E+02	0.130E+02	0.0
10	0.130	0.105E+01	0.913E+01	0.929E+01	0.0	0.130E+00	0.115E+01	0.115E+02	0.150E+02	0.0
11	0.145	0.114E+01	0.104E+02	0.111E+02	0.0	0.145E+00	0.117E+01	0.129E+02	0.171E+02	0.0
12	0.159	0.254E-01	0.104E+02	0.111E+02	0.0	0.159E+00	0.119E+01	0.143E+02	0.193E+02	0.0
13	0.174	0.525E-02	0.104E+02	0.111E+02	0.0	0.174E+00	0.121E+01	0.158E+02	0.217E+02	0.0
14	0.188	0.201E+01	0.145E+02	0.145E+02	0.0	0.188E+00	0.123E+01	0.173E+02	0.242E+02	0.0

15	0.203	0.122E+01	0.160E+02	0.158E+02	0.0	0.203E+00	0.125E+01	0.188E+02	0.255E+02	0.0
16	0.217	0.805E+00	0.166E+02	0.172E+02	0.0	0.217E+00	0.127E+01	0.205E+02	0.268E+02	0.0
17	0.232	0.230E+00	0.167E+02	0.176E+02	0.0	0.232E+00	0.130E+01	0.221E+02	0.283E+02	0.0
18	0.246	0.970E+00	0.176E+02	0.177E+02	0.0	0.246E+00	0.132E+01	0.239E+02	0.298E+02	0.0
19	0.261	0.251E+00	0.177E+02	0.176E+02	0.0	0.261E+00	0.135E+01	0.257E+02	0.314E+02	0.0
20	0.275	0.113E+01	0.190E+02	0.190E+02	0.0	0.275E+00	0.137E+01	0.276E+02	0.332E+02	0.0
21	0.290	0.129E+01	0.206E+02	0.210E+02	0.0	0.290E+00	0.140E+01	0.295E+02	0.350E+02	0.0
22	0.304	0.925E-01	0.206E+02	0.208E+02	0.0	0.304E+00	0.143E+01	0.316E+02	0.370E+02	0.0
23	0.319	0.324E-01	0.206E+02	0.208E+02	0.0	0.319E+00	0.145E+01	0.337E+02	0.390E+02	0.0
24	0.333	0.204E+01	0.248E+02	0.237E+02	0.0	0.333E+00	0.148E+01	0.359E+02	0.412E+02	0.0
25	0.348	0.142E+01	0.268E+02	0.257E+02	0.0	0.348E+00	0.151E+01	0.382E+02	0.435E+02	0.0
26	0.362	0.595E+00	0.273E+02	0.268E+02	0.0	0.362E+00	0.154E+01	0.405E+02	0.460E+02	0.0
27	0.377	0.358E+00	0.274E+02	0.274E+02	0.0	0.377E+00	0.157E+01	0.430E+02	0.486E+02	0.0
28	0.391	0.924E+00	0.283E+02	0.286E+02	0.0	0.391E+00	0.160E+01	0.456E+02	0.513E+02	0.0
29	0.406	0.405E+00	0.284E+02	0.287E+02	0.0	0.406E+00	0.164E+01	0.483E+02	0.545E+02	0.0
30	0.420	0.113E+01	0.298E+02	0.306E+02	0.0	0.420E+00	0.167E+01	0.511E+02	0.579E+02	0.0
31	0.435	0.143E+01	0.319E+02	0.329E+02	0.0	0.435E+00	0.171E+01	0.540E+02	0.614E+02	0.0
32	0.449	0.202E+00	0.319E+02	0.322E+02	0.0	0.449E+00	0.174E+01	0.570E+02	0.652E+02	0.0
33	0.464	0.997E-01	0.319E+02	0.318E+02	0.0	0.464E+00	0.178E+01	0.602E+02	0.691E+02	0.0
34	0.478	0.211E+01	0.364E+02	0.352E+02	0.0	0.478E+00	0.182E+01	0.635E+02	0.733E+02	0.0
35	0.493	0.154E+01	0.391E+02	0.379E+02	0.0	0.493E+00	0.186E+01	0.669E+02	0.777E+02	0.0
36	0.507	0.502E+00	0.394E+02	0.389E+02	0.0	0.507E+00	0.190E+01	0.705E+02	0.824E+02	0.0
37	0.522	0.498E+00	0.397E+02	0.401E+02	0.0	0.522E+00	0.194E+01	0.742E+02	0.873E+02	0.0
38	0.536	0.850E+00	0.404E+02	0.402E+02	0.0	0.536E+00	0.198E+01	0.782E+02	0.925E+02	0.0
39	0.551	0.551E+00	0.407E+02	0.400E+02	0.0	0.551E+00	0.202E+01	0.822E+02	0.979E+02	0.0
40	0.565	0.125E+01	0.423E+02	0.425E+02	0.0	0.565E+00	0.207E+01	0.865E+02	0.104E+03	0.0
41	0.580	0.153E+01	0.448E+02	0.461E+02	0.0	0.580E+00	0.211E+01	0.910E+02	0.110E+03	0.0
42	0.594	0.353E+00	0.449E+02	0.455E+02	0.0	0.594E+00	0.216E+01	0.956E+02	0.116E+03	0.0
43	0.609	0.225E+00	0.450E+02	0.453E+02	0.0	0.609E+00	0.220E+01	0.100E+03	0.119E+03	0.0
44	0.623	0.224E+01	0.500E+02	0.499E+02	0.0	0.623E+00	0.225E+01	0.106E+03	0.123E+03	0.0
45	0.638	0.189E+01	0.536E+02	0.534E+02	0.0	0.638E+00	0.230E+01	0.111E+03	0.126E+03	0.0
46	0.652	0.521E+00	0.538E+02	0.536E+02	0.0	0.652E+00	0.235E+01	0.116E+03	0.130E+03	0.0

47	0.667	0.618E+00	0.542E+02	0.546E+02	0.0	0.667E+00	0.241E+01	0.122E+03	0.134E+03	0.0
48	0.681	0.777E+00	0.548E+02	0.545E+02	0.0	0.681E+00	0.246E+01	0.128E+03	0.138E+03	0.0
49	0.696	0.595E+00	0.553E+02	0.549E+02	0.0	0.696E+00	0.252E+01	0.135E+03	0.142E+03	0.0
50	0.710	0.131E+01	0.570E+02	0.574E+02	0.0	0.710E+00	0.257E+01	0.141E+03	0.146E+03	0.0
51	0.725	0.172E+01	0.600E+02	0.608E+02	0.0	0.725E+00	0.263E+01	0.148E+03	0.150E+03	0.0
52	0.739	0.545E+00	0.603E+02	0.613E+02	0.0	0.739E+00	0.269E+01	0.155E+03	0.155E+03	0.0
53	0.754	0.423E+00	0.605E+02	0.619E+02	0.0	0.754E+00	0.275E+01	0.163E+03	0.159E+03	0.0
54	0.768	0.245E+01	0.665E+02	0.664E+02	0.0	0.768E+00	0.281E+01	0.171E+03	0.164E+03	0.0
55	0.783	0.219E+01	0.713E+02	0.703E+02	0.0	0.783E+00	0.287E+01	0.179E+03	0.169E+03	0.0
56	0.797	0.451E+00	0.715E+02	0.707E+02	0.0	0.797E+00	0.294E+01	0.188E+03	0.174E+03	0.0
57	0.812	0.725E+00	0.720E+02	0.712E+02	0.0	0.812E+00	0.300E+01	0.197E+03	0.181E+03	0.0
58	0.826	0.678E+00	0.725E+02	0.720E+02	0.0	0.826E+00	0.307E+01	0.206E+03	0.188E+03	0.0
59	0.841	0.841E+00	0.732E+02	0.734E+02	0.0	0.841E+00	0.314E+01	0.216E+03	0.196E+03	0.0
60	0.855	0.135E+01	0.750E+02	0.747E+02	0.0	0.855E+00	0.321E+01	0.226E+03	0.205E+03	0.0
61	0.870	0.137E+01	0.785E+02	0.782E+02	0.0	0.870E+00	0.328E+01	0.237E+03	0.214E+03	0.0
62	0.884	0.782E+00	0.791E+02	0.798E+02	0.0	0.884E+00	0.336E+01	0.248E+03	0.224E+03	0.0
63	0.899	0.725E+00	0.797E+02	0.811E+02	0.0	0.899E+00	0.343E+01	0.260E+03	0.234E+03	0.0
64	0.913	0.275E+01	0.873E+02	0.872E+02	0.0	0.913E+00	0.351E+01	0.272E+03	0.246E+03	0.0
65	0.928	0.253E+01	0.937E+02	0.922E+02	0.0	0.928E+00	0.359E+01	0.285E+03	0.257E+03	0.0
66	0.942	0.390E+00	0.938E+02	0.929E+02	0.0	0.942E+00	0.367E+01	0.299E+03	0.270E+03	0.0
67	0.957	0.817E+00	0.945E+02	0.954E+02	0.0	0.957E+00	0.375E+01	0.313E+03	0.284E+03	0.0
68	0.971	0.554E+00	0.948E+02	0.956E+02	0.0	0.971E+00	0.383E+01	0.327E+03	0.298E+03	0.0
69	0.986	0.935E+00	0.958E+02	0.970E+02	0.0	0.986E+00	0.391E+01	0.343E+03	0.313E+03	0.0
70	1.000	0.141E+01	0.978E+02	0.963E+02	0.0	0.100E+01	0.400E+01	0.359E+03	0.306E+03	0.0

A D D E N D A

On va exposer d'une manière sommaire une application du modèle à la resolusion de certaines equations differentielles.

Soit T un operateur : $y \to x = T(y)$ défini par l'equation differentielle :

$$x(t) = \frac{dy}{dt} - f(t,y)$$

avec y et y' des fonctions continues, f(t,y) une fonction continue de t et une fonction entière de y pour chaque valeur de t avec f(t,0)=0.

La solution de cette equation s'écrit formellement :

$$y(t) = T^{-1}[x(t)] = F[x(t)]$$

La methode de resolution proposée consiste à approximer F par un polynome H-S; elle ressemble,en quelque sorte, à celle qui utilise les developpements en série pour résoudre certaines equations différentielles (Bessel,...), la difference etant la representation de la solution par un developpement d'operateurs :

$$y(t) = \sum_n \int_0^t \cdots \int h_n(t,\tau_1,..,\tau_n)x(\tau_1)..x(\tau_n)d\tau_1..d\tau_n$$

au lieu d'un develeppement en série.

La methode a été utilisée [30],[31],[34] sur des equations differentielles du type :

$$L[y(t)] + F[t,y(t),y'(t),..,y^{(n-1)}(t)] = x(t)$$

où L est un operateur differentiel linéaire à coefficients variables, et F un polynome en $y(t),y'(t),..,y^{(n-1)}(t)$; les coefficients de ce polynome sont des fonctions continues de t; n est le plus ordre des derivées figurant dans L et x(t) une fonction continue.

Nous allons donner quelques details sur la determination d'une solution de cette equation avec des conditions initiales nulles.

Pour t=0, y et ses (n-1) derivées sont nulles

On represente la solution par la série :

$$y(t)=\int_0^t h_1(t,\tau)x(\tau)d\tau + \iint_0^t h_2(t,\tau_1,\tau_2)x(\tau_1)x(\tau_2)d\tau_1 d\tau_2 + \ldots$$

$$+ \int_0^t \ldots \int h_n(t,\tau_1,..,\tau_n)x(\tau_1)..x(\tau_n)d\tau_1..d\tau_n + \ldots$$

et on substitue dans l'equation differentielle.

Les noyaux sont alors obtenus en egalant les coefficients des diverses fonctionnelles de x(t).

Nous allons etudier un exemple.

Remarque :

- fonction de Green 'd'un coté'- [32]

Soit L un operateur differentiel linéaire du n^{e} ordre dont les coefficients $p_i(t)$ sont continus dans $[a,\infty)$; une fonction $H(t,\tau)$ est dite fonction de Green d'un coté si :

- $H(t,\tau)$ et ses n premières derivées sont continues par rapport à (t,τ)dans tout fermé de $[a,\infty)$
- $LH(t,\tau)=0$
- $\left.\frac{\partial^\alpha H(t,\tau)}{\partial t^\alpha}\right|_{t=\tau} = 0$ pour $\alpha=0,1,..,n-2$
- $\left.\frac{\partial^{n-1} H(t,\tau)}{\partial t^{n-1}}\right|_{t=\tau} = \frac{1}{p_0(\tau)}$ $p_0$etant le coefficient de $\frac{d^n}{dt^n}$ dans L

On montre que H est unique et que :

si $u(t)=\int_0^t H(t,\tau)f(\tau)d\tau$, alors $Lu=f(t)$ et $u^{(\alpha)}(a)=0$ $\alpha=0,1,..,n-1$.

H etant connu, un système fondamental de solutions de Lu = 0, (n solutions independantes) peut etre écrit explicitement, en fonction de H et de ses derivées à l'extremité a.

H est déterminé à partir de sa définition, des conditions aux limites et de la représentation de la solution u. (on dira fonction de Green à un coté à cause de la détermination de H à partir de l'extrémité a seulement).

Pour la construction de H, les théorèmes sur son unicité ou sur la résolution de Lu=f en fonction de H, on se reportera à [32].

Exemple :

$$A(t)\frac{dy}{dt} + B(t)y + Ky^2 = x(t) \qquad (2)$$

Le calcul sera formel. En utilisant la dérivation sous le signe $\int$ et en identifiant les coefficients des diverses fonctionnelles de x(t) aprés substitution de (1) dans (2), on obtient les equations qui suivent:

fonctionnelles	coefficients	
$x(t)$	$A(t)h_1(t,t)=1$	(3)
$\int_0^t x(\tau)d\tau$	$A(t)\frac{\partial}{\partial t}h_1(t,\tau) + B(t)h_1(t,\tau)= 0$	(4)
$x(t)\int_0^t x(\tau)d\tau$	$A(t)\left\|h_1(t,\tau,t)+h_2(t,t,\tau)\right\|= 0$	(5)
$\int\int_0^t x(\tau_1)x(\tau_2)d\tau_1 d\tau_2$	$A(t)\frac{\partial}{\partial t}h_2(t,\tau_1,\tau_2)+B(t)h_2(t,\tau_1,\tau_2)$ $+ Kh_1(t,\tau_1)h_2(t,\tau_2)= 0$	(6)
. . . . .		
$x(t)\int_0^t..\int \prod_{\substack{i=1\\ i=j}}^{n} x(\tau_i)d\tau_i$	$A(t)\sum_{j=1}^{n} h_n(t,\tau_1,.,\tau_{j-1},t,\tau_{j+1},.,\tau_n)$	(7)
$\int_0^t..\int \prod_{i=1}^{n} x(\tau_i)d\tau_i$	$A(t)\frac{\partial}{\partial t}h_n(t,\tau_1,.,\tau_n)+B(t)h_n(t,\tau_1,.,\tau_n)$ $+ K\sum_{j=1}^{n-1} h_j(\tau_1,.,\tau_j)h_{n-j}(\tau_{j+1},.,\tau_n)= 0$	(8)
. . . . .		

Ces equations peuvent etre résolues et les noyaux obtenus recursivement.
Les equations (3) et (4) donnent :

$$h_1(t,\tau) = \frac{1}{A(\theta)} e^{-\int_{\tau}^{t} \frac{B(\theta)}{A(\theta)} d\theta} \quad \text{pour } t>\tau$$

$$= 0 \quad \text{pour } t<\tau \qquad (9)$$

$h_2(t,\tau_1,\tau_2)$ est déterminé à partir de (6) et de la condition aux limites définie par l'equation (5).

L'equation (6) diffère de (5) seulement par un second membre $-Kh_1(t,\tau_1)h_1(t,\tau_2)$; l'opérateur differentiel des deux equations est le même; elles admettent même fonction de Green. D'apres (4) on peut prendre $h_1(t,\tau)$ comme fonction de Green (d'un coté).

Or, $h_1(t,\tau)$ a déjà été déterminé, d'où l'expression de la solution en terme de fonction de Green :

$$h_2(t,\tau_1,\tau_2) = -K \int_0^t G(t,\theta)h_1(\theta,\tau_1)h_1(0,\tau_2)d\theta$$

avec $G(t,\tau)=h_1(t,\tau)$ pour $t>\tau$

Ceci est la solution de l'equation (5) exprimée à l'aide du noyau h_1 déterminé antérieurement, et satisfaisant les C.L. définies par l'equation (5).

Le noyau h_n est déterminé en termede noyaux d'ordre inferieur à n, à partir des equations (7) et (8).
Ainsi :

$$h_n(t,\tau_1,..,\tau_n)=-K \int_0^t G(t,\theta)\left[\sum_{j=1}^{n-1} h_j(\theta,\tau_1,..,\tau_j)h_{n-j}(\theta,\tau_{j+1},..,\tau_n)\right]$$

pour n=2,3,...

etc...

HISTORIQUE DU PROBLEME

Les premières études sur les systèmes ont eu pour base la representation de ceux-ci par des fonctionnelles et non par des operateurs; à l'entrée $[x(s) \quad s \leqslant t]$, il correspond à l'instant t la sortie y_t; à un instant t fixé cette correspondance associe bien à une fonction un nombre.

Si t change, la representation obtenue pose un problème car la convergence a été etablie pour des fonctionnelles et non pour des operateurs.

Les premières études sur les fonctionnelles remontent à Volterra [23] et Fréchet [24].

Bien plus tard, Cameron et Martin [25] s'appuyant sur un théorème de Paley et Wiener et sur les propriétés d'orthogonalité des polynomes d'Hermite, établirent un developpement de fonctionnelles non lineaires en fonctionnelles orthogonales du type Hermite.

Les premiers problèmes concernant la determination de systèmes non lineaires furent considérés par Wiener [14].
(les problèmes de representation furent envisagés aussi par Pugashev [26] et Zadeh [27], l'un considérant les systèmes non linéaires pouvant être réduits à une combinaison de systèmes linéaires, et l'autre recherchant une representation integrale d'un certain type.)

Dans son étude sur les chaos homogènes, Wiener rencontra des fonctionnelles polynomes du type :

$$\int_{-\infty}^{t} k(s)dx(s,\alpha) \quad , \quad \int_{-\infty}^{t}\int k(s,\sigma)dx(s,\alpha)dx(\sigma,\alpha) \quad , \ldots.$$

x= processus du mouvement brownien

S'appuyant sur les travaux de Cameron et Martin, Wiener supposa qu'un filtre non linéaire pouvait être representé par une combinaison linéaire de telles fonctionnelles (appellées fonctionnelles de Volterra); il utilisa plutôt des fonctionnelles orthogonales pour la representation, et un bruit blanc Gaussien pour l'identification.

Les methodes d'analyse fonctionnelle furent appliquées à l'étude du cas linéaire par Hsieh [17] et Balakrishnan [28]; ce dernier en fit une certaine extension au cas non linéaire [29].

Cependant, le problème de la validité de la representation n'a pas été considéré, en dehors des études qu' on vient de citer sur les fonctionnelles.

Les integrales de convolution multidimensionnelles du modèle peuvent etre obtenues en considérant des systèmes non linéaires simples : l'association d'un système linéaire et d'un système non liéaire sans mémoire (la sortie à l'instant t est déterminée seulement par l'entrée à l'instant t).

Le premier est caracterisé par la reponse impulsionnelle k; il transforme le signal entrée x(t) en un signal de sortie z(t) :

$$z(t) = \int_0^t k(t-\tau)x(\tau)d\tau$$

Le second, caracterisé par une fonction F, transforme z(t) en :

$$y(t) = F[z(t)] = F\left[\int_0^t k(t-\tau)x(\tau)d\tau\right]$$

Si F est continue, le théorème classique de Weiersrass permet de l'approximer par un polynome ($F(z) \sim \Sigma a_i z_i$) , d'où en remplaçant :

$$y(t) = a_0 + a_1 \int_0^t k(t-\tau_1)x(\tau_1)d\tau_1 + a_2 \iint_0^t k(t-\tau_1)k(t-\tau_2)x(\tau_1)x(\tau_2)d\tau_1 d\tau_2$$

$$+ \quad \ldots\ldots$$

ce qui montre que la methode decrite est une généralisation de la représentation des systèmes linéaires.

B I B L I O G R A P H I E

1. Guelfand and Vilenkin : Generalized functions - Volume 4 .
   Application of Harmonic Analysis
   Academic Press.

2. Maurin : Methods of Hilbert spaces
   Warszawa 1967

3. Hille and Phillips : Functional analysis and semi-groups
   American mathematical society
   colloquium publications - Volume 31

4. Prenter : a Weierstrass theorem for real separable Hilbert spaces
   Journal of approximation theory 3. 1970

5. Bourbaki : Integration livre 6 chap.6

6. Loeve : Probability theory University series 1955

7. Guykhman - Skorokhod : Introduction to the theory of random
   processes. Figmatiz 1965

8. Skorokhod : Constructive methods of specifying stochastic
   processes. Russian mathematical surveys
   Vol.30 N.3 May-June 1965

9. Choquet : Mesures affines et cylindriques
   colloque d'analyse fonctionnelle
   Rome Mars 1968

10. L.Schwartz : Applications radonifiantes
    seminaire 1969-70

11. Neveu : Bases mathematiques du calcul des probabilités
    Masson 1964

12. Shilov : Integration in infinite dimensional spaces and the Wiener integral
Russian mathematical surveys 1963 18-1

13. Shilov and Gurevich : Integral,measure and derivative a unified approach
Prentice Hall 1966

14. Wiener : Nonlinear problems in random theory
MIT Press 1958

15. Prosser and Root : Determinables classes of channels
Journal of mathematics and mechanics
Vol.16 N.4 1966

16. Riesz et Nagy : Leçons d'analyse fonctionnelle
Academie de Hongrie

17. Hsieh : The least squares estimation of linear and nonlinear system weighting function matrices
Inf. and Control Vol.7 1964

18. Kichatov : Analytical optimization technique for one class of nonlinear systems.
Automation and remote control August 1965

19. Katznelson and Gould : a spectrum factorization method for the calculation of nonlinear filters of the Volterra type.
Inf. and Control 8 1965

20. Robbins - Monro : a stochastic approximation method
Ann. Math. Statistics 22 N.1 1951

21. Loginov : Methods of stochastic approximation
Survey aut. and remote control Avril 1966

22. Tsypkin : Adaptation and learning in automatic systems
Academic Press 1971

23. Volterra : Théorie generale des fonctionnelles
Gauthiers - Villars 1936

24. Frechet : Sur les fonctionnelles continues
Ann. Ecole Normale 3 - 27 Mai 1910

25. Cameron and Martin : the orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals.
Annals of math. 48 N.2 1947

26. Pugachev : Theory of random functions and its application to control problems. Pergamon Press 1965

27. Zadeh : a contribution to the theory of nonlinear systems
Journal Franklin Inst. 255 1955

28. Balakrishnan : a general theory of nonlinear estimation problems in control systems
Jour. math. analysis and applic. 8 1964

29. Balakrishnan : on the state space theory of linear systems
J. math. analy.appl. 14-3 1965

30. Flake : Volterra series representation of nonlinear systems
IEEE Trans. Appl. Ind. 82-64- 1963

31. Alper : a consideration of the discrete Volterra series
IEEE Tran. Aut. Control AC 10-3-1965

32. Miller : the one-sided Green's function
J. Appl. Phys. 22 1951

33. Aoki : Optimization of stochastic systems
Academic Press 1967

34. Christensen : convergence of Volterra series
IEEE on automatic control dec.1968

35. Christensen and Trott : Initial conditions in Volterra series
International journal of control
Vol.12 N.5 1970

TABLE des MATIERES

Chapitre 0 : Rappels
Opérateurs de Hilbert-Schmidt et opérateurs polynômes. p. 5

PARTIE 1 : Approximation d'opérateurs non linéaires.

Chapitre 1 : Opérateurs non linéaires continus. p. 12

Chapitre 2 : Approximation d'opérateurs non linéaires analytiques. p. 21

PARTIE 2 : Transformations non linéaires d'un processus aléatoire. p. 35

Chapitre 1 : Densité des opérateurs HS dans $L^p(X,\mu)$. p. 36

Chapitre 2 : Transformations non linéaires d'un processus défini par un système projectif de mesures. p. 41

PARTIE 3 : Application à l'identification des systèmes. p. 51

Chapitre 1 : Représentation des systèmes non-linéaires. p. 52

Chapitre 2 : Entrée déterminante pour l'identification. Classes déterminables. p. 58

PARTIE 4 : Identification numérique. p. 71

Chapitre 1 : Relations entre les moments de l'entrée et de la sortie. p. 72

Chapitre 2 : Identification des systèmes non linéaires à entrées-sorties aléatoires. p. 75

Chapitre 3 : Identification des systèmes non linéaires. Cas où les caractéristiques statistiques sont inconnues. p. 80

Chapitre 4 : Approximation Stochastique. p. 84

PARTIE 5 : Simulation.
Algorithmes pratiques, exemples, résultats numériques. p. 95

Addenda : Application du modèle HS à la résolution de certaines équations différentielles. p. 111

--:-:-:--

ACADÉMIE DE PARIS

UNIVERSITÉ DE PARIS VI

4, Place Jussieu - 75230 PARIS-CEDEX 05

SERVICE DE LA SCOLARITE

Bureau des Enseignements du 3e cycle

TÉLÉPHONE : 336-25-25 - 325-12-21

PERMIS D'IMPRIMER

DOCTORAT : . . D'ETAT . . . . . .

THESE DE : . . ES SCIENCES MATHEMATIQUES . . . . . .

NOM : . . ALLAB . . . . . .

Prénoms : . . . Kada . . . . . .

né le 18 Octobre 1942 à . . MASCARA . . . . . Dépt : Algérie

1ère THESE : . . . Approximation d'opérateurs non linéaires et application de l'identification des systèmes . . . . . .

. . . . . .

. . . . . .

. . . . . .

2e THESE (Propositions faites par l'Université) : . . . . . .

. . . . . .

. . . . . .

. . . . . .

Vu et Approuvé

PARIS, le . . 19 Juin 1973

Le Président

de l'Université de PARIS VI

résident: Mr. VILLE
FORTET
LIONS
CHERRUAULT

ACADEMIE DE PARIS
UNIVERSITÉ PARIS VI